Study
Report
2008-02

Longitudinal Junior Noncommissioned Officer Promotion Analysis

Christopher E. Sager
Shaobang Sun
Dan J. Putka
Human Resources Research Organization

Kimberly S. Owens
Tonia S. Heffner
U.S. Army Research Institute

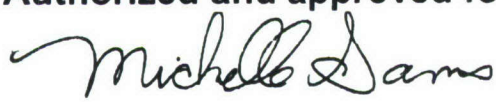
United States Army Research Institute
for the Behavioral and Social Sciences

October 2007

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:



MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical review by

Richard R. Hoffman, U.S. Army Research Institute
Jennifer Solberg, U.S. Army Research institute

NOTICES

DISTRIBUTION: Primary distribution of this Study Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-MS, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Study Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Study Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
October 2007	Final	February 2005 -May 2006

4. TITLE AND SUBTITLE

Longitudinal Junior Noncommissioned Officer Promotion Analysis

5a. CONTRACT OR GRANT NUMBER
DASW01-03-D-0015

5b. PROGRAM ELEMENT NUMBER
665803

6. AUTHOR(S)
Christopher E. Sager, Shaobang Sun, Dan J. Putka (Human Resources Research Organization); Kimberly S. Owens, & Tonia S. Heffner (U.S. Army Research Institute)

5c. PROJECT NUMBER
D730

5d. TASK NUMBER
319

5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Human Resources Research Organization
66 Canal Center Plaza, Suite 400
Alexandria, Virginia 22314

8. PERFORMING ORGANIZATION REPORT NUMBER
FR-06-54

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Institute for the Behavioral and Social Sciences
ATTN: DAPE-ARI-RS
2511 Jefferson Davis Highway
Arlington, Virginia 22202-3926

10. MONITOR ACRONYM
ARI

11. MONITOR REPORT NUMBER
Study Report 2008-02

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Subject Matter POC: Kimberly Owens

14. ABSTRACT *(Maximum 200 words)*:

The Noncommissioned Officer (NCO) Promotion effort was undertaken to help the U.S. Army prepare noncommissioned officers to meet the needs of the future Army. In the earlier NCO21 project, concurrent validation evidence was collected to support the integration of measures of knowledges, skills, and aptitudes (KSAs) into the promotion system. This report documents the longitudinal validation of these measures. The predictor measures included the Leadership Judgment Exercise, Self-Description Inventory, Information Questionnaire-II, Experience and Activity Record, Work Suitability Inventory, and Personnel File Form. Observed and expected future performance rating scales were used to remotely collect criterion data via the Internet a little more than a year later. An additional criterion measure was whether participating Soldiers had been promoted during the project's research period. This project yielded some evidence supporting the longitudinal validity of the predictor measures and good evidence that the measures can effectively be administered via laptop computers.

15. SUBJECT TERMS
Performance(Human), Aptitude Tests, Noncommissioned Officers, Army Personnel, Manpower, Classification, Personnel Selection, Army Planning, Promotion(Advancement)

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	92	Ellen Kinzer Technical Publication Specialist 703-602-8047

Study Report 2008-02

Longitudinal Junior Noncommissioned Officer Promotion Analysis

Christopher E. Sager
Shaobang Sun
Dan J. Putka
Human Resources Research Organization

Kimberly S. Owens
Tonia S. Heffner
U.S. Army Research Institute

Chief, Selection and Assignment Research Unit
Michael G. Rumsey

U. S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

October 2007

Army Project Number
665803D730

Personnel and Training Analysis Activities

LONGITUDINAL JUNIOR NONCOMMISSIONED OFFICER PROMOTION ANALYSIS

EXECUTIVE SUMMARY

Research Requirement:

To ensure that the U.S. Army has high-quality noncommissioned officers (NCOs) prepared to meet the needs of the future Army, a project was initiated to examine possible improvements to NCO promotion systems for the 21st century. This project culminated in a set of predictor measures called the Leadership Assessment Tool (LAT), supported by concurrent criterion-related validity evidence (that is, scores on the predictors were associated with job performance measures,[that is, supervisor ratings], that were administered simultaneously). Based on these positive results, the current project was conceived with three primary goals. The first was to examine whether the evidence supporting the concurrent criterion-related validity of the predictors would extend to a longitudinal validation setting. That is, one in which predictor measures would predict job performance measures (e.g., job performance ratings) collected some time after the predictors were administered. The longitudinal validation setting more closely resembles the operational context where these predictors would be used to aid in promotion decisions predicting future performance at the next pay grade than does the concurrent validation setting. Another goal of this project was to examine the extent to which it would be efficient to administer the predictor measures via laptop computer instead of via paper-and-pencil. The third goal was to determine whether it would be efficient to collect criterion data (i.e., job performance ratings) via the Internet instead of via paper-and-pencil.

Procedure:

Five measures required validation. Four of these measures were part of the original LAT: (a) the Leadership Judgment Exercise (LeadEx), (b) the Self-Description Inventory (SDI), (c) the Information Questionnaire-II (IQ-II), and (d) the Experience and Activity Record (ExAct). The fifth measure—the Work Suitability Inventory (WSI)—was originally developed for another Army personnel research effort. Additionally, the Personnel File Form was used to collect self-report accomplishment information, which was in turn used to compute a Promotion Point Worksheet score that simulated the current promotion system. These measures were administered via laptop computer to E4 and E5 Soldiers who were (or were close to being) eligible for promotion to the next pay grade. These predictor data were collected from 942 E4 and E5 Soldiers.

A little more than a year after the predictor measures were administered, criterion data collection began. E-mail and the Internet were used to collect two types of job performance ratings from the supervisors of these Soldiers. One type was observed performance ratings that assessed how well Soldiers performed their current jobs. The second type was expected future performance ratings in which supervisors were asked to predict how well their Soldiers would perform in conditions expected to be characteristic of the future Army. Because job performance ratings were collected from such a small number of supervisors (i.e., ratings were collected for

only 64 of the original 942 Soldiers), not all the planned validation analyses using this criterion could be performed, and those that were performed need to be interpreted cautiously. In response to this problem, an additional performance criterion was identified—whether or not the Soldier was promoted during the data collection period. Promotion criterion data were collected for 938 Soldiers. The validity of the predictors was assessed by examining the extent to which scores on the predictors were associated with scores on the job performance ratings and the promotion criterion.

Findings:

This project developed some evidence supporting the longitudinal validity of the predictor measures. However, these results need to be interpreted with caution given the small sample size associated with the job performance ratings criterion and conceptual difficulties with the promotion criterion. Because promotion is based on the operational Promotion Point Worksheet, it is not possible to use the promotion criterion to estimate the extent to which the studied predictors could improve the prediction of performance beyond the current system.

This project also showed that, in this context, collecting data using laptop computers is reasonable psychometrically and probably more efficient compared to paper-and-pencil data collection. However, data collection via e-mail and the Internet was not particularly effective at ensuring sufficient rates of participation.

Utilization and Dissemination of Findings:

These results provide some evidence in support of the construct and longitudinal validity of the predictor measures. The findings also support administration of the LAT measures via computer. However, they also provide evidence that the procedures for eliciting further participation from pre-identified Soldiers via e-mail and the Internet need improvement if they are to be effective. Possible approaches for managing this problem include (a) collecting initial predictor data from a much larger number of participants, (b) sending participation solicitation e-mails to Soldiers from superiors who are organizationally more proximate to each Soldier (e.g., a division or installation commander), and (c) ensuring frequent communication with participants between the predictor and criterion data collections (e.g., a newsletter). Finally, further research in an operational setting is recommended to support the assignment of promotion points in the Army's semi-centralized NCO promotion system based on any of these measures.

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

Page

List of Tables

CONTENTS (Continued)

Page

CONTENTS (Continued)

Page

CONTENTS (Continued)

Page

List of Figures

LONGITUDINAL JUNIOR NONCOMMISSIONED OFFICER PROMOTION ANALYSIS

CHAPTER 1: INTRODUCTION

This report describes the longitudinal criterion-related validation of a set of experimental noncommissioned officer (NCO) tools developed as part of a research program sponsored by the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI). The report is targeted toward a technical audience interested in the psychometric characteristics of the measures in the context of their computerization and a longitudinal validation design. Readers interested in more detail on the development of these measures and their performance in a concurrent validation design should see Knapp et al. (2002) and Knapp, McCloy, and Heffner (2004).

Background

To ensure that the U.S. Army has high-quality noncommissioned officers (NCOs) prepared to meet the needs of the future Army, ARI initiated the project titled *Maximizing the Performance of NonCommissioned Officers for the 21st Century (NCO21)*. This project's goal was to examine possible improvements to NCO promotion systems for the 21st century. It culminated in the development and validation of a set of predictor measures called the Leadership Assessment Tool (LAT). The LAT was designed to improve promotion decisions for specialists/corporals (E4s) and sergeants (E5s) to the next pay grade. The concurrent validation effort showed promising results regarding the construct and predictive validity of the LAT predictors (Knapp et al. 2004). Indeed, there was good evidence for incremental validity beyond the current promotion system. The reasonable inference was made that a predictor demonstrating criterion-related validity in a concurrent setting would likely demonstrate validity in a longitudinal setting that has more fidelity with the operational context. However, concern was expressed about whether the relative contribution of these predictors would remain fixed given their nature. For example, it was acknowledged that performance on some of these predictors is likely influenced by experience and training. This project's primary goal was to investigate the possibility that the validity of the predictors would be different when examined in the longitudinal context. Another goal was to examine the extent to which it would be practical and psychometrically reasonable to collect (a) data on the predictor measures via laptop computer instead of paper-and-pencil and (b) criterion data (i.e., job performance ratings) via e-mail and the Internet instead of paper-and-pencil in a controlled data collection setting.

Phase I of this project was titled the *Leadership Potential Assessment for the Non-Commissioned Officer (NCO) Junior Promotion System Analysis*. Its objectives were to begin a longitudinal validation by collecting predictor data and examining the psychometric characteristics of LAT scores in the context of computer administration compared to the original paper-and-pencil administration of the instruments. Phase II, titled *Longitudinal Junior Noncommissioned Officer Promotion Analysis: Criterion*, focused on collection of criterion data.

Longitudinal Criterion-Related Validation

This project differed from the concurrent validation (Knapp et al., 2004) in three important ways. First, the predictor measures were administered via laptop instead of paper-and-pencil. Second, the criterion job performance ratings from supervisors were collected via e-mail and the Internet instead of in-person using paper-and-pencil instruments. Third and most importantly, this project used a longitudinal validation design (in which the predictors are administered, some period of time passes, and then criteria are administered) instead of a concurrent design (in which the predictor and criterion measures are administered at the same time). The predictor instruments discussed in this report were administered between June and October of 2004; data on the criterion measures were collected between December 2005 and February 2006.

This project, however, was similar to the concurrent validation in an important way, beyond the fact that it used the same measures: Its experimental predictor and criterion measures focused on assessing the knowledges, skills, and aptitudes (KSAs), and behaviors relevant to current and expected future performance. The criterion supervisor ratings included 21 scales designed to assess dimensions of current observed job performance and 6 scales designed to assess performance in future conditions forecasted for NCOs by a future-oriented job analysis (Ford, Knapp, J. Campbell, R. Campbell, and Walker, 2000). The LAT predictors were designed to assess KSAs relevant to current and expected future performance (Knapp et al., 2004).

Predictor Data Collection

The LAT included seven instruments (see Figure 1.1) that were administered by laptop computer to Soldiers during a 4-hour session. The first instrument, the Soldier Background Information Form (SBIF), is not a predictor. It collects basic personal identifying and demographic information (e.g., name, project identification number, location, pay grade, and Military Occupational Specialty (MOS), Army Knowledge Online [AKO] e-mail address). The first of the predictor measures is the Personnel File Form (PFF21). It is used to collect information for simulating current promotion system selection factors (e.g., Awards, Military Education, Military Training, and Civilian Education). The Leadership Judgment Exercise (LeadEx) is a situational judgment test designed to assess Soldiers' judgments about potential courses of action in response to job-related scenarios. The Self-Description Inventory (SDI) and the Information Questionnaire-II (IQ-II) are operational temperament measures used in the Army for other purposes (Kilcullen, Chen, Zazanis, Carpenter, & Goodwin, 1999; Kilcullen, Mael, Goodwin, & Zazanis, 1999; Kilcullen, White, Zacarro, & Parker, 2000; White & Young, 1998; Young, Heggestad, Rumsey, & White, 2000). The IQ-II is actually a compilation of multiple measures. The experimental versions of both the SDI and IQ-II used here were prepared for the original NCO21 project (Knapp et al., 2004). The Experience and Activities Record (ExAct) queries Soldiers about work experiences, activities, and accomplishments not directly assessed in the current promotion system. The Work Suitability Inventory (WSI) is an experimental measure designed to assess temperament constructs related to work. It was developed during another ARI project (i.e., Select21; McCloy & Putka, 2005) and was not originally part of the LAT. The LeadEx and ExAct are experimental measures that were developed specifically for the original NCO21 project. Additional data were collected from the Enlisted Master File (EMF) including race/ethnicity, gender, and General Technical (GT) scores from the Armed Services Vocational

Aptitude Battery (ASVAB). These data were accessed using the social security numbers (SSNs) of Soldiers in the predictor database and matching them with Soldier SSNs in the archival database.

Order	Instrument
1.	Soldier Background Information Form (SBIF)
2.	Personnel File Form-21 (PFF21)
3.	Leadership Judgment Exercise (LeadEx)
4.	Self-Description Inventory (SDI)
5.	Information Questionnaire-II (IQ-II)
6.	Experience and Activities Record (ExAct)
7.	Work Suitability Inventory (WSI)

Figure 1.1 Leadership Assessment Tool (LAT) instruments.

Criterion Data Collection

The criterion data collection procedure consisted first of Soldiers, who participated in the predictor data collection, logging on to the NCO Promotion Soldier website and (a) nominating supervisors who could rate their job performance, (b) providing some demographic information, (c) completing the same PFF21 from the predictor data collection with some additional items asking about the Soldier's latest promotion and promotion system scores, and (d) completing the same ExAct from the predictor data collection. Additional data, including information about current pay grade, time in service (TIS), and time in grade (TIG) were collected from the Enlisted Master File (EMF) and Military Enlistment Processing Command Integrated Resource System (MIRS). The second part of the criterion data collection required each nominated supervisor to log on to the NCO Promotion Supervisor website and provide current observed and expected future job performance ratings of the Soldier or Soldiers who nominated that supervisor.

Overview of Report

Chapter 1 discussed the background, goals, and general structure of the data collections for this project. Chapter 2 presents the method and additional details of the predictor and criterion data collections such as sample sizes at each stage of the data collection, and details regarding data cleaning and database development. Chapter 3 describes the psychometric characteristics of each instrument administered during the predictor data collection. Chapter 4 does the same for instruments administered via the Soldier and Supervisor websites during the criterion data collection. Chapter 5 presents cross-instrument analyses, including relations among predictors and longitudinal criterion-related validity results. Finally, Chapter 6 summarizes the findings of this research.

CHAPTER 2: DATA COLLECTION AND DATABASE DEVELOPMENT

Introduction

This chapter describes the longitudinal validation data collection, sample sizes at each stage of the research, construction of the analysis database, and administration times for the predictor measures. The predictor dataset, after data cleaning, included 591 E4 and 351 E5 Soldiers. During the first part of the criterion data collection, 73 E4 and 69 E5 Soldiers logged on to the NCO Promotion Soldier website to nominate supervisor raters and complete the criterion data collection versions of the PFF21 and ExAct. During the second part of the criterion data collection 75 supervisors provided ratings for 36 E4 and 28 E5 Soldiers. At the end of the criterion data collection, the MIRS archival database was queried to determine which of the original participants were still in the Army and whether they had been promoted since they had completed the experimental predictor measures. These promotion data were obtained for 588 E4 and 350 E5 Soldiers.

Predictor Data Collection Procedures

Between June and October of 2004, data were collected from E4 and E5 Soldiers near eligibility for promotion to the next pay grade. A two-step process determined whether Soldiers were near eligibility for promotion. First, Soldiers were included if they were within 9 months of the time in service (TIS) and time in grade (TIG) requirements for promotion to the next grade (i.e., 27 months TIS and no TIG requirement for E4 Soldiers and 75 months TIS and 1 month TIG for E5 Soldiers). Second, if Soldiers were not eligible in this way, they were asked if they had received a wavier to be eligible for early promotion. If the answer was yes, E4 Soldiers still needed at least 12 months TIS and E5 Soldiers needed at least 42 months TIS. Soldiers who were not eligible were dismissed before the data collection session began.

E4 and E5 Soldiers were scheduled for a 4-hour session during which the seven instruments described in Chapter 1 were administered via laptop computer. For each instrument, if the Soldier failed to respond to an item, he/she was reminded of the missing data and was afforded a second chance to provide the missing information. If the missing data were not provided the second time, the software moved on to the next item.

As part of the administrative procedure, a 2 x 2 between-subjects design varied two factors: (a) instrument order for the LeadEx, SDI, and IQ-II and (b) item order for these instruments. Figure 2.1 illustrates the resulting four instrument administration conditions. Each laptop computer was labeled and included the software to support only one of these conditions. Individuals were assigned to laptops such that during each session roughly an equal number of Soldiers completed the LAT under each condition. These two administration factors were varied across participants to control for and assess carryover effects (e.g., fatigue) for these relatively long instruments and their items. The "instrument order" factor was limited to two levels (i.e., the LeadEx before and after the other two instruments) because the primary concern was that the amount of reading required of Soldiers to complete the LeadEx would produce carryover effects that would negatively affect their performance on the SDI and IQ-II. Appendix A provides internal consistency reliability and mean score results showing that instrument and item order had very little effect.

Condition	Factor 1: Instrument Order	Factor 2: Item Order
1	LeadEx, SDI, and IQ-II	Original order used in concurrent validation data collection
2	SDI, IQ-II, LeadEx	Original order used in concurrent validation data collection
3	LeadEx, SDI, and IQ-II	Second half of the items first; first half second[1]
4	SDI, IQ-II, LeadEx	Second half of the items first; first half second

Figure 2.1 Description of predictor administration conditions.

Table 2.1 shows sample sizes following data cleaning procedures for the total sample and key subgroups used in the analyses (e.g., pay grade, gender, and race/ethnicity).[2] After data cleaning, the final sample included 591 E4 and 351 E5 Soldiers. According to military occupational specialty (MOS), the participating Soldiers were sorted into three categories: (a) Combat Arms (CA), (b) Combat Support (CS), and (c) Combat Service Support (CSS). Table 2.1 also presents the number of participants at each of six data collection sites.

Table 2.1. Demographic Composition of Predictor Data Collection Sample

	E4 Soldiers		E5 Soldiers	
Group	*N*	%	*N*	%
Gender				
Male	498	84.3	307	87.5
Female	93	15.7	44	12.5
Race/Ethnicity				
White	344	58.2	217	61.8
Black	123	20.8	95	27.1
Hispanic	74	12.5	28	8.0
Other	48	8.1	11	3.1
MOS Type				
Combat Arms	225	38.1	147	41.9
Combat Support	109	18.4	45	12.8
Combat Service Support	257	43.5	159	45.3
Administration Location				
Fort Campbell	66	11.2	54	15.4
Fort Hood	51	8.6	8	2.3
Fort Lewis	143	24.2	93	26.5
Fort Riley	89	15.1	57	16.2
Fort Sill	169	28.6	60	17.1
Korea	73	12.4	79	22.5

Note. $n_{E4} = 591$. $n_{E5} = 351$. Sample sizes are based on gender, race/ethnicity, and primary MOS data obtained from the December 2004 EMF file. For two Soldiers, values for gender and MOS that they reported on the background form were used because of unavailability of EMF data. Actual analysis sample sizes may be smaller than the totals listed here due to missing or unusable data at the instrument level.

[1] There is a minor exception in the SDI. The very first item is the same because it is an unscored practice item.
[2] The data cleaning procedures are described in the Database Construction and Cleaning section of this chapter.

Criterion Data Collection Procedures

Soliciting Soldier Participation

Figure 2.2 shows the schedule for the criterion data collection. The first e-mail sent to Soldiers was signed by the Chief of the Enlisted Career Systems Division in the Office of the Deputy Chief of Staff, G1. This solicitation e-mail (a) explained the importance of the Soldier's participation, (b) reminded the Soldier of his/her earlier participation in the predictor data collection, and (c) explained that the Soldier would soon receive an e-mail with further instructions and a link to the NCO Promotion Soldier website. This solicitation e-mail was sent to 926 of the original 942 Soldiers who participated in the predictor data collection. After cleaning and correcting e-mail addresses provided by participating Soldiers during the predictor data collections, 865 had a usable Army Knowledge On-line (AKO) address, 569 had an alternate personal address, and 508 had both. Only 16 Soldiers did not provide a usable e-mail address. Table 2.2 shows the number of participants at each stage of the Soldier phase of criterion data collection.

12/01/05	ARI sent Soldier solicitation e-mail
12/06/05	HumRRO sent Soldier participation e-mail
12/13/05	HumRRO sent 1st Soldier reminder e-mail
12/21/05	HumRRO sent 2nd Soldier reminder e-mail
01/04/05	HumRRO sent 3rd and final Soldier reminder e-mail with a January 13, 2006 deadline
01/30/05	ARI sent supervisor solicitation e-mail
02/01/05	HumRRO sent supervisor participation e-mail
02/08/05	HumRRO sent 1st supervisor reminder e-mail
02/15/05	HumRRO sent 2nd and final supervisor reminder e-mail with February 24, 2006 deadline

Figure 2.2 Criterion Data Collection Schedule.

Table 2.2. Soldier Participation in Criterion Data Collection

Stage	*N*	%
Sent solicitation e-mail from ARI	926	
Sent Soldier participation e-mail from HumRRO	926	
Soldiers responding before first reminder	43	4.64%
Soldiers responding between first and second reminder	28	3.02%
Soldiers responding between second and third reminder	27	2.92%
Soldiers responding after third reminder	43	4.64%
Total Soldier respondents	141	15.23%

Note. Two additional Soldiers responded by logging on to the NCO Promotion Soldier Website, but declined to participate further by disagreeing with the Privacy Act Statement.

Shortly after Soldiers received the solicitation e-mail from ARI, they received the participation e-mail from HumRRO. This e-mail contained the following items:

- A reminder of the solicitation e-mail and past participation in the predictor data collection,
- Instructions for nominating supervisors to rate the Soldier's job performance,
- A link to the NCO Promotion Soldier website,

- An individual password for the website,
- Contact information for help, and
- A project briefing.

According to the schedule shown in Figure 2.2, Soldiers who had not yet responded received reminder e-mails. The reminder e-mail consisted of the original participation e-mail, including the Soldier website link and password, preceded by text reminding the Soldier about the previously sent solicitation and participation e-mails and the importance of the Soldier's participation in the research. The original plan included only two reminder e-mails; however, a third reminder was added to the schedule because the second reminder was sent just before Christmas. The original participation e-mail and first two reminders asked the Soldier to respond as soon as possible. The third and final reminder requested that the Soldier respond by January 13, 2006. Table 2.3 shows sample sizes following data cleaning for the Soldiers who provided data on the NCO Promotion Soldier website. The first set of columns represents the Soldiers whose pay grade was E4 when they completed the predictor instruments. The columns labeled E3 through E7 indicate the pay grade of these Soldiers when their archival Army records were queried at the end of this longitudinal analysis (December 31, 2005). The second set of columns shows the same data for Soldiers who were E5s when they completed the predictor instruments. Table 2.3 shows that the majority of Soldiers either stayed at the same pay grade or were promoted once, although a small number were demoted or promoted more than once.

Table 2.3. Demographic Composition of Soldiers Participating via the NCO Promotion Website.

	Pay Grade During Predictor Data Collection												
	E4							E5					
	Pay Grade Reported on Website							Pay Grade Reported on Website					
Group	E3	E4	E5	E6	E7	Total *n*	% Total	E4	E5	E6	E7	Total *n*	% Total
Gender													
Male	1	23	29	3	1	57	78.1	2	33	23	1	59	86.8
Female	0	5	11	0	0	16	21.9	0	7	1	0	9	13.2
Race/Ethnicity													
White	0	16	21	1	0	38	52.1	0	28	12	1	41	60.3
Black	0	9	13	0	1	23	31.5	0	7	6	0	14	20.6
Hispanic	1	2	5	2	0	10	13.7	2	2	5	0	9	13.2
Other	0	1	1	0	0	2	2.7	0	3	1	0	4	5.9
MOS Type													
Combat Arms	1	10	10	0	0	21	28.8	0	12	5	1	18	26.5
Combat Support	0	5	5	0	0	10	13.7	1	9	7	0	17	25.0
Combat Service Support	0	13	25	3	1	42	57.5	1	19	12	0	33	48.5

Note. $n_{E4} = 73$. $n_{E5} = 68$. Sample sizes are based on gender, race/ethnicity, and primary MOS data obtained from the December 2004 EMF file. For two Soldiers, values for gender and MOS that they reported on the background form were used because of unavailability of EMF data. One E5 Soldier did not report current pay grade on the website; therefore the rows for female, black, and combat service support Soldiers are one Soldier short; however, the Total *n*s are correct.

NCO Promotion Soldier Website

The first screen of the NCO Promotion Soldier website required the participant to enter his/her e-mail address and website password. This was followed by an opportunity to review the project briefing that was provided in the participation e-mail. Next, the project's privacy act statement was presented and the Soldier was asked to check a box agreeing with its conditions. If the Soldier disagreed, the information was saved, and the Soldier was logged off the website. If the Soldier agreed, the website moved on to the nomination of two supervisors who could rate the Soldier's job performance. The requirements for eligibility to be a supervisor rater were as follows:

> Supervisors can be NCOs, Warrant Officers, and/or Commissioned Officers. The best choice for your First Supervisor is your direct supervisor (First Line Supervisor). The best choice for your Second Supervisor is your direct supervisor's supervisor (Second Line Supervisor). It is important that your supervisors know you well. If you haven't worked with your direct supervisor or your Second Line Supervisor for at least one (1) month, replace either of them with a superior who has recently observed your performance for one (1) month or more. This alternate supervisor does not have to be someone who supervised you as long as he or she is in a supervisory job.

This text also appeared in the Soldier's participation e-mail. The website asked for the names, AKO and alternate (i.e., personal) e-mail addresses, and work telephone numbers of the nominated supervisors. Throughout the website, if the Soldier failed to provide any of the requested information, he/she was reminded of the missing data and was afforded a second chance to provide the missing information. If the missing data were not provided the second time, the website moved on to the next page.

After the Soldier nominated supervisor raters, the website asked a few demographic questions (i.e., name, location, current MOS). This was followed by some questions about the Soldier's latest promotion and current Promotion Point Worksheet points. Finally, the Soldier completed the same PFF21 and ExAct from the predictor data collection.

Soliciting Supervisor Participation

Figure 2.2 shows the schedule for the criterion data collection. The supervisor solicitation e-mail was sent by ARI and was signed by the Chief of the Enlisted Career Systems Division in the Office of the Deputy Chief of Staff, G1. This solicitation e-mail had the same content as the Soldier solicitation e-mail except that it explained that one of the supervisor's Soldiers had participated in the earlier predictor data collection portion of the NCO Promotion Analysis. After addresses had been cleaned and corrected, 137 of the 141 Soldiers who provided data on the NCO Promotion Soldier website provided at least one usable supervisor e-mail address (123 Soldiers provided addresses for two supervisors; 14 provided addresses for only one supervisor). This resulted in solicitation e-mails being sent to 252 supervisors. Table 2.4 shows the number of supervisor participants and Soldiers for whom ratings were solicited or collected at each stage of this part of the criterion data collection.

Table 2.4. Supervisor Participation in Criterion Data Collection

	Supervisor Raters		Soldier Ratees	
Stage	*N*	%	*n*	%
Sent solicitation e-mail from ARI	252		137	
Sent supervisor participation e-mail from HumRRO	252		137	
Supervisors responding before first reminder	28	11.11%	26[b]	20.47%
Supervisors responding between first and second reminder	25	9.92%	21[b]	15.33%
Supervisors responding after second reminder	22	8.73%	17[b]	12.41%
Total supervisor respondents	75[a]	29.76%	64[b]	46.72%

[a] This total includes six supervisors who indicated that they had not worked with their Soldier for at least a month and therefore were not asked to provide ratings.
[b] These values reflect the number of Soldiers who had received ratings from at least one supervisor at each stage.

All of the 75 supervisors who logged on to the NCO Promotion website agreed to its privacy act statement and 69 moved on to the ratings portion of the website after indicating they had worked with their Soldier for at least a month. These supervisors provided ratings for 64 Soldiers (53 Soldiers were rated by one supervisor, and 11 Soldiers were rated by two supervisors). While this number amounts to a 29.76% response rate for supervisors (see Table 2.4), it amounts to a 46.72% response rate in terms of the percentage of 137 Soldiers who received ratings from at least one supervisor, and an 8.03% response rate in terms of the percentage of Soldiers who received ratings from two supervisors. Table 2.5 shows the demographic characteristics of these Soldiers and supervisors.

Table 2.5. Demographic Composition of Supervisors and Their Soldiers Participating via the NCO Promotion Website.

	Supervisor Raters		Soldier Ratees	
Group	*N*	%	*N*	%
Gender				
Male	66	88.0	49	76.6
Female	9	12.0	15	23.4
Race/Ethnicity				
White	43	57.3	38	59.4
Black	19	25.3	16	25.0
Hispanic	6	8.0	9	14.1
Other	7	9.3	1	1.6
MOS Type				
Combat Arms	18	24.0	19	29.7
Combat Support	16	21.3	17	26.6
Combat Service Support	23	30.7	28	43.8
Warrant/Commissioned Officer	18	24.0		

Note. $n_{Supervisors} = 75$. $n_{Soldiers} = 64$. The supervisor sample sizes are based on gender, race/ethnicity, and primary MOS self-report data obtained from the NCO Promotion Supervisor website. The Soldier values are from the December 2004 EMF. For two Soldiers, values for gender and MOS that they reported on the background form were used because of unavailability of EMF data.

NCO Promotion Supervisor Website

The NCO Promotion Supervisor website began the same as the Soldier website in terms of the password, briefing, and privacy act statement. Supervisors then were asked to provide basic demographic information for themselves (e.g., MOS, pay grade, gender, and race/ethnicity). Next, supervisors were presented with the names of the Soldier(s) who had nominated them and were asked to indicate how long they had worked with the Soldier(s).[3] If they indicated that they had not worked with the Soldier for at least a month, the supervisors were not presented with the rating scales for that Soldier.

After it was determined that the supervisor was eligible to rate a Soldier, the supervisor was presented with instructions for making observed performance ratings. Appendix B shows these instructions including the layout of the observed performance rating scales that were then presented one at a time (see Figure 2.3 for a list of scale titles).[4] Similar to the Soldier website, if the supervisor failed to provide any of the requested information (e.g., a rating on a particular scale), he/she was reminded of the missing data and was afforded a second chance to provide the missing information. If the missing information was not provided the second time, the website moved on to the next page. After the supervisor made ratings on 21 scales (i.e., 19 dimensions of observed performance, one overall effectiveness scale, and one senior NCO potential scale), the supervisor was presented with a complete list of his/her ratings. At this point the supervisor had the opportunity to click on any rating, return to that rating scale, and change the rating. Next, the supervisor was asked to evaluate his/her ratings on a 7-point confidence scale.

1. MOS/Occupation-Specific Knowledge and Skill
2. Common Task Knowledge and Skill
3. Computer Skills
4. Writing Skill
5. Oral Communication Skill
6. Level of Effort/Initiative on the Job
7. Adaptability
8. Self-Management and Self-Directed Learning Skill
9. Demonstrated Integrity, Discipline, and Adherence to Army Procedures
10. Acting as a Role Model
11. Relating to and Supporting Peers
12. Cultural Tolerance
13. Selfless Service Orientation
14. Leadership Skills
15. Concern for Soldier Quality of Life
16. Training Others
17. Coordinating Multiple Units and Battlefield Functions
18. Problem-Solving/Decision Making Skill
19. Information Management
20. Overall Effectiveness
21. Senior NCO Potential

Figure 2.3 Titles of Observed Performance Rating Scales.

[3] Most supervisors were nominated by only one Soldier; however, the website was developed to accommodate up to five Soldiers per supervisor.

[4] The complete text of the observed performance rating scales is in Knapp, McCloy, and Heffner (2004).

After the observed performance ratings were made, the supervisor examined four pages of briefing slides describing anticipated future conditions that NCOs are likely to face in the future Army. These conditions were based on a future-oriented job analysis reported by Ford et al. (2000). The briefing was provided to help supervisor raters understand the difference between observed performance and expected future performance.

The briefing was followed by a set of rating instructions and six expected future performance rating scales, presented one at a time (see Figure 2.4 for a list scale titles).[5] Appendix B shows the instructions and the first expected future performance rating scale. Similar to the observed performance ratings, supervisors were reminded once if they did not make a rating on a particular scale. After the supervisor made ratings on the six scales, the supervisor was presented with a complete list of his/her ratings. At this point the rater had the opportunity to click on any rating, return to that rating scale, and change the rating. Finally, the next page asked the supervisor to evaluate his/her rating on each expected future performance scale using a 7-point confidence scale.

1. Increased Requirements for Self-Direction and Self-Management
2. Use of Computers, Computerized Equipment, and Digitized Operations
3. Increased Scope of Technical Skill Requirements
4. Increased Requirements for Broader Leadership Skills at Lower Levels
5. Need to Manage Multiple Operational Functions and Deal with the Inter-relatedness of Units
6. Mental and Physical Adaptability and Stamina

Figure 2.4 Titles of Expected Future Performance Rating Scales.

An Alternative Criterion

As can be seen by the sample sizes discussed in Tables 2.1, 2.2, and 2.4, participation diminished substantially at each stage. The predictor data collection included 942 Soldiers (i.e., 591 E4 and 351 E5 Soldiers). Only 141 Soldiers logged on to the NCO Promotion Soldier website and agreed to participate (i.e., 15.0% of the original 942). This low response rate resulted in potential supervisor raters being contacted for only 137 Soldiers. Participation was solicited from 252 supervisors, 75 of whom logged on to the NCO Promotion Supervisor website, resulting in at least one supervisor rater for each of only 64 Soldiers. This meant that, before data cleaning, there were criterion data in the form of job performance ratings for only 6.8% of the original sample (i.e., [64/942]100).

Faced with this difficulty, we sought to develop an alternative criterion to the job performance ratings. The criterion we selected was whether or not participants had been promoted by the time the criterion data collection was completed. The MIRS database was queried to identify the most recent promotion and pay grade for each of the 942 Soldiers who had participated in the original predictor data collection as of December 31, 2005 (i.e., the end of the criterion data collection).

To use promotion as a criterion in the analyses, another variable needed to be created. It is referred to here as "exposure" and reflects an estimate of the number of months a Soldier had been eligible to be promoted at the time the criterion data collection ended (i.e., December 31,

[5] The complete text of the expected future performance rating scales is in Knapp, McCloy, and Heffner (2004).

2005) or at the time the Soldier left the Army, whichever came first. The exposure variable was developed for two reasons. First, a validity analysis that uses promotion as the criterion should include only those Soldiers who had at least some minimal opportunity to be promoted in terms of exposure. The value of 6 months was selected as a reasonable minimal period. Second, exposure itself could be a predictor of promotion. For example, up to a certain number of months of exposure, the relation between exposure and promotion could be positive after which it could turn negative (i.e., additional exposure could result in a reduced probability of promotion). The use of exposure in the validation analyses is discussed in Chapter 5.

The following values were used to calculate exposure for each Soldier: (a) self-report TIS, (b) the standard policy that E4 Soldiers need 27 months TIS and E5 Soldiers need 75 months TIS to be eligible for promotion to the next pay grade, (c) the end date for criterion data collection (i.e., December 31, 2005), and (d) separation dates for Soldiers who left the service before the end of the data collection (obtained from the MIRS database). After eliminating Soldiers who had missing MIRS data, unrealistic self-reported TIS values, and/or an exposure value of less than 6 months, this data set included 513 E4 and 260 E5 Soldiers.

Database Construction and Cleaning

Predictor Data Collection

Several steps were taken to ensure the quality of the data collected. First, the Soldier paper rosters that included the name, pay grade, computer identification number, participant identification number, and administration condition for each Soldier were compared to the same information collected on the laptops to ensure its accuracy. Second, information from session logs was used to identify and eliminate response from Soldiers with questionable data. Third, for the ExAct, LeadEx, SDI, and IQ-II data, Soldiers who failed to respond to at least 90% of the items were dropped from further analyses. No Soldiers were dropped for missing data on the PFF21 because the items required participants to endorse achievements (e.g., medals, awards, and letters of commendation). If an item was left blank, the Soldier simply did not get credit for that accomplishment. The Missing Data columns in Table 2.6 reflect Soldiers who were dropped from further analyses because (a) their data for that instrument were identified as questionable in a session log or (b) they responded to fewer than 90% of the items. The WSI is a special case; it is constructed such that none of the responses are recorded unless the participant responds to all of the items. Therefore, Missing Data values for this instrument represent the number of Soldiers who did not complete the WSI. These relatively large numbers are not surprising given that the WSI was the last instrument administered and Soldiers occasionally exited from the administration software without responding to all of its items. With the exception of the WSI, the relatively small amount of missing data was expected given that the predictor administration software generated a warning every time a Soldier advanced to the next item without responding to the current item. Next, because the computer software collected precise individual test administration times, we were able to drop the scores of participants who completed an instrument so quickly that their responses could not be an accurate reflection of their standing on the constructs being assessed (see the Testing Time column in Table 2.6 for these losses). These data suggest that hurrying through an instrument was a more common phenomenon among E4 Soldiers than among E5 Soldiers. Finally, ExAct, LeadEx, SDI, and IQ-II responses were screened for patterned or illogical response patterns. For example, we looked for Soldiers who repeatedly gave the same response to too many items or gave the same response to adjacent items so infrequently that they might have been pattern responding. Table 2.6 shows that data for very few Soldiers were

eliminated for pattern responding. After completing the data cleaning steps illustrated in Table 2.6, the remaining data were sufficiently complete that we determined that imputation of missing data was not necessary.

Table 2.6. Predictor Sample Sizes by Instrument and Data Cleaning Results

	E4 Soldiers					E5 Soldiers				
			Reason for Data Loss					Reason for Data Loss		
Instrument	Usable *n*	% Loss	Missing Data	Testing Time	Response Pattern	Usable *n*	% Loss	Missing Data	Testing Time	Response Pattern
PFF21	591	0.0	0	0	0	351	0.0	0	0	0
ExAct	574	2.9	5	12	0	347	1.1	3	1	0
LeadEx	551	6.8	4	36	0	339	3.4	1	11	0
SDI	581	1.7	2	6	2	346	1.4	2	1	2
IQ-II	579	2.0	3	4	5	345	1.7	2	2	2
WSI	540	8.6	11	40	0	322	8.3	19	10	0

Note. $n_{E4} = 591$. $n_{E5} = 351$. Usable *n* = number of Soldiers with usable data for the given instrument. % Loss = percentage of Soldiers in the overall sample whose data were deemed unusable. Missing data = number of Soldiers who failed to respond to at least 90% of the instrument's items. Testing Time = number of Soldiers who completed the instrument in an unreasonably short time. Response Pattern = number of Soldiers who exhibited patterned responding on the instrument (among Soldiers whose data were not lost due to missing items or testing time). Actual analysis sample sizes may be smaller than the usable sample sizes listed here due to missing data at the scale-level.

After data cleaning, scale scores were calculated for each instrument. Scores were calculated for all Soldiers on all PFF21 scales based on the accomplishments they endorsed. Consistent with item formats and operational scoring of this instrument, Soldiers did not get points for awards, training, and other accomplishments that they did not affirmatively indicate that they earned. Because one 24-item and one 40-item composite LeadEx score was calculated for each Soldier, individuals remaining after cleaning were assigned these two scores based on their averages across the relevant completed items.[6] The ExAct, SDI, and IQ-II generate multiple scale scores. For each scale, a minimum number of necessary items per scale was identified. If a Soldier completed this minimum, a score was calculated based on the average across the completed items. This procedure resulted in no missing scale scores for the SDI and IQ-II and only one missing score for one ExAct scale. As described above, scores from the WSI were not recorded until the Soldier completed all items; therefore it had no missing scale scores.

Criterion Data Collection

Data from the Soldier Website

The steps for preparing the data collected during this stage were similar to those followed during the predictor data collection. For the criterion version of the PFF21 and ExAct, Soldiers who failed to respond to at least 90% of the items were dropped from further analyses. The Missing Data columns in Table 2.7 reflect Soldiers who were dropped from further analyses

[6] One LeadEx score was based on the all 40 items included in the experimental version of this instrument. Another LeadEx score was based on a subset of 24 items identified during the concurrent validity project as optimal candidates for an operational length version of this instrument.

because they responded to fewer than 90% of the items or because they demonstrated patterned or illogical response patterns. Although the number of individuals dropped because of missing data was small, the relative percentage was greater than comparable values resulting from the predictor data collection. Only one Soldier on one instrument was eliminated for pattern responding. No scores were dropped for testing times that were too short.

Table 2.7. Soldier Criterion Sample Sizes by Instrument and Data Cleaning Results

	E4 Soldiers				E5 Soldiers			
			Reasons for Data Loss				Reasons for Data Loss	
Instrument	Usable *n*	% Loss	Missing Data	Response Pattern	Usable *n*	% Loss	Missing Data	Response Pattern
PFF21	71	2.74	2	0	66	2.94	2	0
ExAct	70	4.11	3	0	65	4.41	2	1

Note. $n_{E4} = 73$. $n_{E5} = 68$. Usable *n* = number of Soldiers with usable data for the given instrument. % Loss = percentage of Soldiers in the overall sample whose data were deemed unusable. Missing data = number of Soldiers who failed to respond to at least 90% of the instrument's items. Testing Time = number of Soldiers who completed the instrument in an unreasonably short time. Response Pattern = number of Soldiers who exhibited patterned responding on the instrument (among Soldiers whose data were not lost due to missing items or testing time). Actual analysis sample sizes may be smaller than the usable sample sizes listed here due to missing data at the scale-level.

As with the predictor data collection, we determined that imputation of missing data was not necessary. Again, this relatively small amount of missing data was expected given that the NCO Promotion Soldier website software generated a warning every time a Soldier advanced to the next item without responding to the current item. Scales on the PFF21 and the ExAct were scored in the same manner as they were in the predictor data collection.

Data from the Supervisor Website

The steps for preparing data collected during this stage were similar to those followed during the predictor and Soldier portion of the criterion data collection. First, a supervisor's ratings of a Soldier were dropped if the supervisor reported working with the Soldier for less than one month. Table 2.8 shows that six sets of supervisor ratings were lost for this reason. Next, for the observed performance rating scales, the ratings for supervisors who failed to respond to at least 90% of the items were dropped from further analyses.[7] The Missing Data column in Table 2.8 reflects supervisors who were dropped from further analyses because they responded to fewer than 90% of the items. The same procedure was followed for the six expected future performance scales. Next, responses were screened for patterned responding or completion times that were too short. No supervisor's ratings were eliminated for pattern responding or for completing the ratings too quickly. As with the other data collections, we determined that imputation of missing data was not necessary. For each ratee, the Observed Performance Composite score was calculated based on the mean of the Observed Performance scales that were rated. The Expected Future Performance Composite score was calculated the same way.

[7] Out of 21 scales (i.e., 19 observed performance scales, 1 overall effectiveness scale, and 1 senior NCO potential scale, Scale 17 (Coordinating Multiple Units and Battlefield Functions) was eliminated from this and further analyses because of the low rate response rate for that scale (26.1% of the supervisor raters indicated that they could not rate their Soldier on this scale). This value was 22.8% in the concurrent validation (Sager, Putka, & McCloy, 2004).

Table 2.8. Supervisor and Soldier Ratings Sample Sizes after Data Cleaning

	Supervisor Raters				
			Reasons for Data Loss		Usable Soldier
Instrument	Usable *n*	% Loss	< 1 Month	Missing Data	Ratee *n*
Observed Composite	66	12.0	6	3	56[a]
Expected Future Composite	61	18.7		8	53

Note. Usable *n* = number of supervisors with usable data for the given instrument. % Loss = percentage of supervisors in the overall sample whose data were deemed unusable. < 1 Month = number of supervisors who didn't work with their Soldiers long enough to rate their performance. Missing data = number of supervisors with too many missing ratings. Usable Soldier Ratee *n* = resulting number of Soldiers with at least one usable set of ratings.
[a] The number of usable supervisor ratings does not agree with the number of usable Soldier ratings because some Soldiers were rated by two supervisors.

Administration Times

Table 2.9 shows test administration times for the predictor data collection instruments administered on laptops. These administration times compare favorably to the estimated administration times for the paper-and-pencil versions used during the concurrent validation data collection. However, it is important to note that paper administration times are much less precise because they are estimates, whereas the times for the laptop computer administration are actual times recorded by the computer program for each individual participant. The paper-and-pencil values were the prescribed amount of time for administration of each instrument to a group. Focusing on administration times at the 90th percentile, Table 2.9 suggests time savings for the PFF21, ExAct, and LeadEx, but not for the SDI or the IQ-II.

Table 2.9. Time Statistics for Predictor Data Collection by Instrument (in minutes)

	Concurrent Validation Paper Administration	Longitudinal Validation Computer Administration Time							
		E4 Soldiers				E5 Soldiers			
Instrument	Time	*Mdn*	*SD*	90th Percentile	95th Percentile	*Mdn*	*SD*	90th Percentile	95th Percentile
PFF21	20	5.1	2.1	8.1	9.7	5.3	5.0	8.3	10.0
ExAct	15	7.2	1.8	9.5	10.4	7.4	1.6	9.5	10.2
LeadEx	65	34.4	11.5	51.9	58.3	35.3	17.4	53.9	62.2
SDI	30	19.6	6.7	28.5	31.9	19.7	7.0	29.8	33.9
IQ-II	40	28.6	7.8	40.5	45.3	27.7	7.8	39.7	44.3
WSI	[a]	4.1	8.6	6.6	7.8	4.2	6.4	6.9	8.0

Note. $n_{Concurrent}$ = 1,881-1891. $n_{Longitudinal, E4}$ = 540-591. $n_{Longitudinal, E5}$ = 322-351. Statistics are based on Soldiers with usable instrument data. *Mdn* = number of minutes by which 50% of Soldiers completed the instrument. *SD* = standard deviation of instrument completion times. 90% Percentile = number of minutes by which 90% of Soldiers completed the instrument. 95% Percentile = number of minutes by which 95% of Soldiers completed the instrument.
[a] The WSI was not administered during the concurrent validation.

Table 2.10 shows criterion data collection administration times for the two instruments that were also administered during the predictor data collection. The median times for completing these instruments on the website versus laptop computers (see Table 2.9) are very

similar. The difference is that the 90th and 95th percentile times are longer for the website administration. This finding is not a surprise, given that during the predictor data collection Soldiers were responding to the instruments in a relatively quiet "testing" environment. Website administration occurred at a computer that the Soldier chose and could have included a number of interruptions and/or distractions.

Table 2.10. Time Statistics for the Soldier Website Data Collection by Instrument (in minutes)

	E4 Soldiers				E5 Soldiers			
Instrument	*Mdn*	*SD*	90th Percentile	95th Percentile	*Mdn*	*SD*	90th Percentile	95th Percentile
PFF21	4.9	3.7	10.1	13.5	5.5	5.6	11.0	15.9
ExAct	7.0	3.1	13.1	13.9	7.2	4.3	13.3	15.9

Note. n_{E4} = 70-71. n_{E5} = 65-66. Statistics are based on Soldiers with usable instrument data. *Mdn* = number of minutes by which 50% of Soldiers completed the instrument. *SD* = standard deviation of instrument completion times. 90% Percentile = number of minutes by which 90% of Soldiers completed the instrument. 95% Percentile = number of minutes by which 95% of Soldiers completed the instrument.

Summary

This chapter described the NCO Promotion Analysis longitudinal validation data collection effort and procedures for processing and cleaning the data. Participants included E4 and E5 Soldiers who were, or were close to, being eligible for promotion to the next pay grade when they completed a number of experimental predictors. Between 14 and 19 months later, Soldiers logged on to the NCO Promotion Soldier website and nominated supervisors who could rate their job performance. Soon after, the nominated supervisors logged on to the NCO Promotion Supervisor website and rated the job performance of their Soldiers. The remaining chapters present and discuss analyses of the resulting data.

CHAPTER 3: RESULTS FOR PREDICTOR DATA COLLECTION INSTRUMENTS

Overview

This chapter documents the results of analyses conducted for each predictor instrument administered during the predictor data collection. Given the salience of pay grade differences found in the NCO21 concurrent validation effort (see Knapp et al., 2004), all results are presented by pay grade. For each instrument, we provide results regarding:

- Mean score differences across pay grades,
- Internal consistency reliability estimates (where appropriate),
- Correlations among instrument scales, and
- Mean score differences across demographic subgroups (gender, race/ethnicity, and MOS).

Simulated Promotion Point Worksheet (SimPPW)

The *operational* Promotion Point Worksheet (PPW) forms the basis of the Army's current NCO promotion system at the E5 and E6 levels. Soldiers receive promotion points in six areas on the operational PPW: (a) Commander's Evaluation; (b) Promotion Board points; (c) Awards, Certificates, and Military Achievements; (d) Military Education; (e) Civilian Education; and (f) Military Training. Promotion points for the first two areas are awarded by a Soldier's commander and promotion board members at the time a Soldier is up for promotion, whereas points for the latter four areas are allocated by the personnel system based on Soldier records. The *simulated* PPW (SimPPW) was developed as part of a broader instrument called the Personnel File Form-21 (PFF21). The PFF21 was designed as a self-report measure for capturing Soldiers' operational PPW data in the latter four areas.[8] Details on the development of the PFF21 are presented in Knapp et al. (2002).

Scoring of the SimPPW

Given that the SimPPW was administered as part of the NCO21 concurrent validation effort, we only briefly describe the scales that constitute the instrument. A more complete description of these scales is available in Putka and Campbell (2004).

SimPPW Awards

The operational PPW credits Soldiers with promotion points for obtaining various awards, certificates, and military achievements. A simulated PPW Awards score was calculated by assigning promotion points to self-reported awards, certificates, and military achievements (based on operational PPW point specifications) and summing these points for each Soldier.[9] SimPPW Award scores were capped at 100 points to be consistent with operational practice.

---

[8] Reasons for exclusion of the first two promotion point areas are discussed in Knapp et al. (2004).

[9] Promotion point specifications are based on AR 600-8-19: Enlisted Promotions and Reductions (Department of the Army, 2004).

SimPPW Military Education

The operational PPW also gives Soldiers promotion points for completing various military education programs. A simulated PPW Military Education score was calculated by assigning promotion points to self-reported military educational experiences (again, based on operational PPW point specifications) and summing these points for each Soldier. SimPPW Military Education scores were capped at 200 points to be consistent with operational practice.

SimPPW Military Training

The operational PPW gives Soldiers promotion points for achieving high levels of marksmanship and physical fitness. A simulated PPW Military Training score was calculated by assigning promotion points to self-reported Army Physical Fitness Test (APFT) and weapons qualification scores (based on operational PPW point specifications) and summing these points for each Soldier.[10] SimPPW Military Training scores were capped at 100 points to be consistent with operational practice.

SimPPW Civilian Education

The operational PPW gives Soldiers promotion points for completing various types of civilian higher education. A simulated PPW Civilian Education score was calculated by assigning promotion points to self-reported civilian educational experiences (based on operational PPW point specifications) and summing these points for each Soldier. SimPPW Civilian Education scores were capped at 100 points to be consistent with operational practice.

SimPPW Composite

A simulated PPW Composite score was calculated for each Soldier by summing the four simulated scores described above. The maximum score that a Soldier could receive on this composite was 500. Note that this maximum score differs from the maximum score on the operational PPW because the simulated PPW does not include Commander's Evaluation points (150) or Promotion Board points (150).

SimPPW Scores by Pay Grade

Table 3.1 shows mean SimPPW scores by pay grade. Like the NCO21 concurrent validation sample, E5 Soldiers were found to have much higher SimPPW scores than E4 Soldiers, particularly with regard to Awards, Military Education, and the overall composite. In comparison to the concurrent validation sample, we found that Soldiers in this sample tended to have lower Military Education scores (particularly among E5 Soldiers, $M_{LV} = 39.30$ vs. $M_{CV} = 63.09$) but similar scores on other scales.

---

[10] A recent change to the operational PPW resulted in a more complicated method for assigning promotion points for weapons qualification. As in the concurrent validation effort, here we used the simpler original promotion point assignment method (e.g., Unqualified = 0, Marksman = 10) because of expectations about what Soldiers could accurately remember.

Table 3.1. Mean SimPPW Scores by Pay Grade

Scale	$d_{E5\text{-}E4}$	E4 Soldiers		E5 Soldiers	
		M	*SD*	*M*	*SD*
Awards	**1.49**	45.93	26.96	86.19	20.77
Military Education	**1.44**	10.53	20.04	39.30	36.41
Military Training	**0.54**	49.92	21.39	61.57	21.69
Civilian Education	**0.75**	11.27	24.00	29.20	36.61
SimPPW Composite	**1.87**	117.65	52.70	216.26	63.75

Note. $n_{E4} = 591$, $n_{E5} = 351$. $d_{E5\text{-}E4}$ = effect size for E5-E4 mean difference. Effect sizes calculated as $(M_{E5} - M_{E4})/SD_{E4}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 3.2 shows correlations among SimPPW scores by pay grade. These results are similar to those found in the concurrent validation sample.

Table 3.2. SimPPW Scale Intercorrelations

Scale	1	2	3	4
E4 Soldiers				
1. Awards				
2. Military Education	**.16**			
3. Military Training	**.13**	**.15**		
4. Civilian Education	.04	**.08**	.05	
5. SimPPW Composite	**.64**	**.56**	**.55**	**.52**
E5 Soldiers				
1. Awards				
2. Military Education	**.14**			
3. Military Training	.02	-.03		
4. Civilian Education	.08	**.11**	-.06	
5. SimPPW Composite	**.46**	**.67**	**.30**	**.64**

Note. $n_{E4} = 591$. $n_{E5} = 351$. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

SimPPW Scores by Gender

Table 3.3 shows mean SimPPW scores by gender for Soldiers in each pay grade. As in the concurrent validation sample, male Soldiers tended to have higher Military Training scores and lower Civilian Education scores than female Soldiers. Differences in Military Training were more pronounced for E4 Soldiers, whereas differences in Civilian Education were more pronounced for E5 Soldiers. Unlike the concurrent validation sample, we found that female E5 Soldiers scored significantly higher on Military Education and the SimPPW composite compared to male E5 Soldiers. In contrast, no significant differences were found between females and males on Military Education and the SimPPW composite at the E4 pay grade.

Table 3.3 Mean SimPPW Scores by Gender

Scale	d_{F-M}	Male		Female	
		M	*SD*	*M*	*SD*
E4 Soldiers					
Awards	**-0.26**	47.03	26.99	40.06	26.19
Military Education	0.03	10.43	20.07	11.12	19.95
Military Training	**-0.57**	51.82	21.33	39.74	18.77
Civilian Education	**0.33**	10.07	23.05	17.67	27.86
SimPPW Composite	-0.21	119.34	50.09	108.59	64.51
E5 Soldiers					
Awards	0.11	85.91	20.63	88.18	21.84
Military Education	**0.47**	37.28	34.35	53.45	46.43
Military Training	-0.29	62.34	21.27	56.18	23.99
Civilian Education	**0.57**	26.67	35.36	46.82	40.64
SimPPW Composite	**0.53**	212.20	61.77	244.64	70.59

Note. $n_{\text{Male E4}} = 498$, $n_{\text{Female E4}} = 93$. $n_{\text{Male E5}} = 307$, $n_{\text{Female E5}} = 44$. d_{F-M} = effect size for Female-Male mean difference. Effect sizes calculated within pay grade as $(M_{\text{Female}} - M_{\text{Male}})/SD_{\text{Male}}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Differences in effect sizes across validation samples can be traced back to differences *within* genders across validation samples. For example, the finding of significant gender differences on Military Education among E5 Soldiers arises from the fact that male E5 Soldiers in the longitudinal sample tended to have notably lower Military Education scores ($M_{LV} = 37.28$) than male E5 Soldiers in the concurrent sample ($M_{CV} = 62.11$). A similar, but smaller trend was seen in Military Education scores for female E5 Soldiers ($M_{LV} = 53.45$; $M_{CV} = 68.18$). Similarly, the finding of significant gender differences on the SimPPW composite among E5 Soldiers arises from the fact that male E5 Soldiers in the longitudinal sample had notably lower SimPPW scores ($M_{LV} = 212.20$) than male E5 Soldiers in the concurrent sample ($M_{CV} = 233.32$). Conversely, female E4 Soldiers in the longitudinal sample tended to have higher SimPPW scores ($M_{LV} = 244.64$) than female E5 Soldiers in the concurrent sample ($M_{CV} = 237.86$). One potential explanation for these findings is that a greater proportion of male Soldiers in the longitudinal sample may have been deployed (compared to male Soldiers in the concurrent sample) and, as such, may have had reduced opportunities for military education.

SimPPW Scores by Race/Ethnicity

Table 3.4 shows mean SimPPW scores by race/ethnicity for Soldiers in each pay grade. No significant differences were found between Whites and Hispanics on any of the SimPPW scores. Additionally, Black-White differences also were quite small, as only one effect size exceeded 0.30. The overall pattern of results was quite similar to those found in the concurrent validation sample. For example, minimal Black-White differences were found with regard to Awards, and Black E5 Soldiers had higher Military Education scores than did White E5 Soldiers. Unlike the concurrent validation sample, we found a small (yet significant) race difference on SimPPW composite scores for E4 Soldiers in this sample (i.e., Blacks scored higher than Whites).

Table 3.4. Mean SimPPW Scores by Race/Ethnic Group

Scale	$d_{B\text{-}W}$	$d_{H\text{-}W}$	White		Black		Hispanic	
			M	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers								
Awards	0.13	-0.03	45.46	26.36	48.76	28.72	44.74	28.97
Military Education	0.11	0.13	9.84	20.15	12.04	20.86	12.47	21.70
Military Training	-0.01	-0.09	50.06	21.08	49.92	23.09	48.19	21.73
Civilian Education	**0.26**	0.15	9.38	22.67	15.19	28.65	12.69	22.94
SimPPW Composite	**0.22**	0.07	114.74	50.26	125.91	60.98	118.09	54.72
E5 Soldiers								
Awards	0.18	-0.15	85.37	21.26	89.15	19.03	82.25	23.64
Military Education	**0.41**	-0.02	35.60	31.46	48.60	46.70	34.82	25.51
Military Training	**-0.29**	-0.13	63.60	21.71	57.34	21.35	60.82	21.45
Civilian Education	0.17	-0.19	28.31	35.54	34.40	39.59	21.71	32.43
SimPPW Composite	**0.28**	-0.22	212.88	59.95	229.48	71.68	199.61	56.33

Note. $n_{White\ E4} = 344$, $n_{Black\ E4} = 123$. $n_{Hispanic\ E4} = 74$, $n_{White\ E5} = 217$, $n_{Black\ E5} = 95$. $n_{Hispanic\ E5} = 28$. $d_{B\text{-}W}$ = effect size for Black-White mean difference. $d_{H\text{-}W}$ = effect size for Hispanic-White mean difference. Effect sizes calculated within pay grade as (mean of non-referent group – M_{White})/SD_{White}. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

SimPPW Scores by MOS

Table 3.5 shows mean SimPPW scores by MOS type for Soldiers in each pay grade. Examination of Table 3.5 reveals that the largest differences were found for E5 Soldiers in CSS MOS. With the exception of the Military Training score, E5 CSS Soldiers had significantly higher

Table 3.5. Mean SimPPW Scores by MOS Type

Scale	$d_{CS\text{-}CA}$	$d_{CSS\text{-}CA}$	$d_{CSS\text{-}CS}$	Combat Arms		Combat Support		Combat Service Support	
				M	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers									
PPW Awards	-0.18	0.10	**0.28**	45.61	26.03	40.76	28.30	48.40	26.97
PPW Military Education	-0.06	**0.29**	**0.34**	8.26	15.63	7.10	11.86	13.98	25.11
PPW Military Training	**-0.43**	**-0.35**	0.08	54.90	20.87	45.64	19.90	47.37	21.69
PPW Civilian Education	0.01	0.12	0.12	9.99	24.40	10.11	20.58	12.88	24.97
Simulated PPW Composite	**-0.29**	0.07	**0.36**	118.76	46.08	103.61	49.76	122.64	58.21
E5 Soldiers									
PPW Awards	-0.13	**0.42**	**0.56**	82.56	22.12	79.78	24.75	91.37	16.79
PPW Military Education	0.22	**0.49**	0.26	30.24	19.74	38.36	39.34	47.96	44.65
PPW Military Training	**-0.50**	**-0.44**	0.06	67.27	20.96	56.38	19.77	57.77	21.78
PPW Civilian Education	**0.40**	**0.82**	**0.42**	13.72	25.26	28.29	36.24	43.77	39.71
Simulated PPW Composite	0.14	**0.74**	**0.60**	193.78	45.44	202.80	62.00	240.86	69.95

Note. $n_{CA\ E4} = 225$, $n_{CS\ E4} = 109$. $n_{CSS\ E4} = 257$, $n_{CA\ E5} = 147$, $n_{CS\ E5} = 45$. $n_{CSS\ E5} = 159$. $d_{CS\text{-}CA}$ = effect size for Combat Support-Combat Arms mean difference. $d_{CSS\text{-}CA}$ = effect size for Combat Service Support-Combat Arms mean difference. $d_{CS\text{-}CA}$ = effect size for Combat Support-Combat Service Support mean difference. Effect sizes calculated within pay grade as (mean of 1st MOS type – mean of 2nd MOS type)/Overall *SD*. Overall *SD* = standard deviation calculated across all Soldiers in the given pay grade (regardless of MOS type). Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

SimPPW scores than did E5 Soldiers in CA MOS. The finding of elevated SimPPW composite scores among E5 CSS Soldiers is consistent with results from the concurrent validation sample. In that sample, Soldiers in "administrative" Career Management Fields (CMF) had notably higher SimPPW composite scores than Soldiers in other CMF. Not surprisingly, we found Soldiers in CA MOS had significantly higher Military Training scores than Soldiers in other MOS.

Experience and Activities Record (ExAct)

The ExAct is a 46-item self-report measure designed to capture information about Soldiers' work experiences, activities, and accomplishments that are indicative of knowledge, skills, and attributes (KSAs) relevant to the performance of 21st-century NCOs (Ford et al., 2000). The content of the ExAct reflects specific activities and experiences that are not typically documented but might predict performance at the next pay grade. It is a reasonable presumption that Soldiers who have engaged in a greater number of these activities and have engaged in them more frequently, often will perform at a higher level than will Soldiers with less experience. That is, knowledge of a Soldier's prior experiences should provide useful information for assessing his or her preparedness to perform similar activities in the future. Details on the development of the ExAct can be found in Knapp et al. (2002).

Scoring of the ExAct

In the concurrent validation effort, we found that a three-factor solution adequately accounted for the covariation among the ExAct items (Putka, 2004). Based on results of these factor analyses, we formed three ExAct scale scores for use in subsequent validation analyses (i.e., Computer Experience, Supervisory Experience, and General Experience). For the present research, we adopted the same scoring algorithm that was used in the concurrent validation effort.

ExAct Scores by Pay Grade

Table 3.6 shows mean ExAct scores by pay grade. Like the concurrent validation sample, E5 Soldiers were found to have much higher Supervisory and General Experience scores than E4 Soldiers. Overall, Soldiers in this sample had higher Supervisory and General Experience scores (particularly among E4 Soldiers) and similar Computer Experience scores compared to those in the concurrent validation sample. For example, E4 Soldiers in this sample had mean Supervisory and General Experience scores of -0.35 and -0.24 respectively, whereas E4 Soldiers in the concurrent validation sample had mean Supervisory and General Experience scores of -0.95 and -0.59, respectively.

Caution should be taken in interpreting mean differences between these samples due to the fact that ExAct scores were standardized within each sample, and the composition of the samples differed. Specifically, E4 Soldiers constituted roughly 62% of the longitudinal validation sample but only about 24% of the concurrent validation sample. Also, unlike the longitudinal sample, the concurrent sample included E6 Soldiers. In fact, E6 Soldiers accounted for roughly 30% of the concurrent sample. Given the differences in Supervisory and General Experience found between pay grades within each sample and differences in sample composition, it is clear that Soldiers in the concurrent sample had a higher mean experience level compared to Soldiers in the longitudinal

sample. Standardizing experience scores within samples likely masks this mean difference and makes it appear that Soldiers in the longitudinal sample have higher experience scores than Soldiers in the concurrent sample. In other words, standardizing within each sample creates a situation where experience scores of "0" do not correspond to the same level of experience in each sample.

Table 3.6. Mean ExAct Scores by Pay Grade

		E4 Soldiers		E5 Soldiers	
Scale	$d_{E5\text{-}E4}$	*M*	*SD*	*M*	*SD*
Computer Experience	**0.31**	-0.08	0.67	0.13	0.56
Supervisory Experience	**1.45**	-0.35	0.65	0.59	0.41
General Experience	**1.48**	-0.24	0.43	0.40	0.41

Note. n_{E4} = 573-574, n_{E5} = 347. $d_{E5\text{-}E4}$ = effect size for E5-E4 mean difference. Effect sizes calculated as $(M_{E5} - M_{E4})/SD_{E4}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 3.7 shows correlations among ExAct scores by pay grade, as well as internal consistency reliability estimates for each scale. For the most part, these results are similar to those found in the concurrent validation sample. The primary difference is that the correlation between Computer and Supervisory Experience among E4 Soldiers is far higher in this sample (.30) than it was in the concurrent validation sample (.06).

Table 3.7. ExAct Scale Intercorrelations and Reliability Estimates

Scale	1	2	3
E4 Soldiers			
1. Computer Experience	(.81)		
2. Supervisory Experience	.30	(.89)	
3. General Experience	.31	.70	(.86)
E5 Soldiers			
1. Computer Experience	(.75)		
2. Supervisory Experience	.13	(.82)	
3. General Experience	.24	.56	(.81)

Note. n_{E4} = 572-573. n_{E5} = 347. Internal consistency reliability estimates (alpha) are shown in parentheses on the diagonal. All correlations are statistically significant, $p < .05$ (one-tailed).

ExAct Scores by Gender

Table 3.8 shows mean ExAct scores by gender for Soldiers in each pay grade. Significant gender differences were found on all of the ExAct scales, with all differences being more pronounced for E5 Soldiers compared to E4 Soldiers. The pattern of differences was identical to the pattern found in the concurrent validation sample: male Soldiers scored higher on Supervisory and General Experience, whereas female Soldiers scored higher on Computer Experience. The magnitudes of these gender differences were also similar to those found in the concurrent validation effort.

Table 3.8. Mean ExAct Scores by Gender

Scale	$d_{F\text{-}M}$	Male		Female	
		M	*SD*	*M*	*SD*
E4 Soldiers					
Computer Experience	**0.22**	-0.10	0.68	0.05	0.62
Supervisory Experience	**-0.27**	-0.32	0.66	-0.50	0.57
General Experience	**-0.42**	-0.21	0.44	-0.40	0.35
E5 Soldiers					
Computer Experience	**0.43**	0.10	0.57	0.35	0.43
Supervisory Experience	**-0.58**	0.62	0.39	0.39	0.45
General Experience	**-0.59**	0.43	0.40	0.19	0.42

Note. $n_{\text{Male E4}}$ = 481-482, $n_{\text{Female E4}}$ = 92. $n_{\text{Male E5}}$ = 304, $n_{\text{Female E5}}$ = 43. $d_{F\text{-}M}$ = effect size for Female-Male mean difference. Effect sizes calculated within pay grade as $(M_{\text{Female}} - M_{\text{Male}})/SD_{\text{Male}}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

ExAct Scores by Race/Ethnicity

Table 3.9 shows mean ExAct scores by race/ethnicity for Soldiers in each pay grade. No significant differences were found between Whites and Blacks for any of the ExAct scores, and only one significant difference was found between Whites and Hispanics (among E4 Soldiers, Hispanics scored lower than Whites on General Experience). The finding of minimal race differences on the ExAct is consistent with results of the concurrent validation effort, where only one statistically significant race difference was found (E4 White-Black on General Experience).

Table 3.9. Mean ExAct Scores by Race/Ethnic Group

Scale	$d_{B\text{-}W}$	$d_{H\text{-}W}$	White		Black		Hispanic	
			M	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers								
Computer Experience	-0.08	-0.09	-0.05	0.67	-0.11	0.69	-0.11	0.61
Supervisory Experience	0.07	-0.13	-0.34	0.64	-0.29	0.75	-0.42	0.61
General Experience	0.01	**-0.33**	-0.22	0.43	-0.22	0.47	-0.36	0.36
E5 Soldiers								
Computer Experience	-0.13	-0.06	0.16	0.53	0.09	0.63	0.13	0.56
Supervisory Experience	0.03	0.01	0.59	0.37	0.60	0.48	0.60	0.41
General Experience	-0.20	-0.29	0.44	0.41	0.36	0.37	0.32	0.44

Note. $n_{\text{White E4}}$ = 340, $n_{\text{Black E4}}$ = 116-117. $n_{\text{Hispanic E4}}$ = 70-71, $n_{\text{White E5}}$ = 214, $n_{\text{Black E5}}$ = 94. $n_{\text{Hispanic E5}}$ = 28. $d_{B\text{-}W}$ = effect size for Black-White mean difference. $d_{H\text{-}W}$ = effect size for Hispanic-White mean difference. Effect sizes calculated within pay grade as (mean of non-referent group – $M_{\text{White}})/SD_{\text{White}}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

ExAct Scores by MOS

Table 3.10 shows mean ExAct scores by MOS type for Soldiers in each pay grade. Among E4 Soldiers, two significant MOS differences were found, and both regarded Computer Experience. Specifically, E4 Soldiers in Combat Service (CS) MOS had significantly higher Computer Experience scores than did E4 Soldiers in other MOS. Among E5 Soldiers, several significant differences emerged. For the most part these differences involved Soldiers in CA

MOS. Specifically, E5 Soldiers in CA MOS had significantly lower Computer Experience scores, and significantly higher Supervisory Experience scores compared to E5 Soldiers in other MOS. These results are similar to those in the concurrent validation sample.

Table 3.10. Mean ExAct Scores by MOS Type

				Combat Arms		Combat Support		Combat Service Support	
Scale	$d_{CS\text{-}CA}$	$d_{CSS\text{-}CA}$	$d_{CSS\text{-}CS}$	*M*	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers									
ExAct Computer Experience	**0.55**	-0.05	**-0.60**	-0.13	0.68	0.24	0.61	-0.16	0.65
ExAct Supervisory Experience	0.04	0.01	-0.03	-0.36	0.62	-0.33	0.61	-0.35	0.69
ExAct General Experience	-0.01	-0.11	-0.10	-0.22	0.42	-0.22	0.41	-0.27	0.45
E5 Soldiers									
ExAct Computer Experience	**0.40**	**0.35**	-0.05	0.01	0.57	0.24	0.60	0.21	0.51
ExAct Supervisory Experience	**-0.51**	**-0.30**	0.22	0.68	0.35	0.47	0.44	0.55	0.43
ExAct General Experience	-0.19	**-0.29**	-0.10	0.46	0.41	0.39	0.37	0.35	0.41

Note. $n_{CA\ E4}$ = 215-216, $n_{CS\ E4}$ = 107. $n_{CSS\ E4}$ = 250-251, $n_{CA\ E5}$ = 144, $n_{CS\ E5}$ = 45. $n_{CSS\ E5}$ = 158. $d_{CS\text{-}CA}$ = effect size for Combat Support-Combat Arms mean difference. $d_{CSS\text{-}CA}$ = effect size for Combat Service Support-Combat Arms mean difference. $d_{CS\text{-}CA}$ = effect size for Combat Support-Combat Service Support mean difference. Effect sizes calculated within pay grade as (mean of 1st MOS type – mean of 2nd MOS type)/Overall *SD*. Overall *SD* = standard deviation calculated across all Soldiers in the given pay grade (regardless of MOS type). Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Leadership Judgment Exercise (LeadEx)[11]

The LeadEx is a 40-item situational judgment test. Situational judgment tests assess the effectiveness of examinees' judgments about the appropriate courses of action in various job-related scenarios. Each item on the LeadEx presents Soldiers with a 2–4 sentence scenario (i.e., description of a problem situation) followed by four possible actions. Soldiers are instructed to indicate (a) which action would be *most* effective, and (b) which action would be *least* effective. The LeadEx was designed to tap eight of the NCO21 KSAs (Ford et al., 2000), with five items representing each KSA. A detailed description of the development of the LeadEx is provided in Knapp et al. (2002).

Scoring of the LeadEx

Consistent with the concurrent validation effort, two LeadEx composite scores are examined here: one based on 24 items, the other based on all 40 items. The scoring of both LeadEx composites is based upon subject matter experts' (SMEs') ratings of the effectiveness of response options used in each item (see Knapp et al., 2002). The score for each LeadEx item is computed by subtracting the keyed effectiveness (i.e., the SMEs' mean effectiveness rating) of the option selected by the Soldier as *least* effective from the keyed effectiveness of the option selected as *most* effective. The LeadEx composite scores were formed by taking the mean across the resulting item scores. Further details on scoring the LeadEx, as well as a discussion of all of the scoring options originally considered for the LeadEx, are presented in Waugh (2004).

[11] Note that in previous project reports, the LeadEx was called the Situational Judgment Test (SJT).

LeadEx Scores by Pay Grade

Table 3.11 shows mean LeadEx scores by pay grade. As in the concurrent validation sample, E5 Soldiers were found to have LeadEx scores that were roughly one-half standard deviation higher than E4 Soldiers.

Table 3.11. Mean LeadEx Scores by Pay Grade

		E4 Soldiers		E5 Soldiers	
Composite	$d_{E5\text{-}E4}$	*M*	*SD*	*M*	*SD*
40-Item	**0.54**	1.87	0.61	2.19	0.47
24-Item	**0.52**	1.89	0.63	2.21	0.52

Note. $n_{E4} = 551$, $n_{E5} = 339$. $d_{E5\text{-}E4}$ = effect size for E5-E4 mean difference. Effect sizes calculated as $(M_{E5} - M_{E4})/SD_{E4}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 3.12 shows correlations among LeadEx scores by pay grade, as well as internal consistency reliability estimates for each composite. Both LeadEx scores exhibited adequate levels of internal consistency, and these estimates were comparable to those observed in the concurrent validation sample.

Table 3.12. LeadEx Scale Intercorrelations and Reliabilities

Composite	1	2
E4 Soldiers		
1. 40-Item	(.82)	
2. 24-Item	.94	(.72)
E5 Soldiers		
1. 40-Item	(.77)	
2. 24-Item	.93	(.69)

Note. $n_{E4} = 551$. $n_{E5} = 339$. Internal consistency reliability estimates (alpha) are shown in parentheses on the diagonal. All correlations are statistically significant, $p < .05$ (one-tailed).

LeadEx Scores by Gender

Table 3.13 shows mean LeadEx scores by gender for Soldiers in each pay grade. Like the concurrent validation sample, female E4 Soldiers had significantly higher LeadEx scores than male E4 Soldiers, and no significant gender differences were found among E5 Soldiers.

Table 3.13. Mean LeadEx Scores by Gender

		Male		Female	
Composite	$d_{F\text{-}M}$	*M*	*SD*	*M*	*SD*
E4 Soldiers					
40-Item	**0.33**	1.84	0.60	2.03	0.61
24-Item	**0.28**	1.86	0.62	2.03	0.65
E5 Soldiers					
40-Item	-0.16	2.20	0.46	2.13	0.55
24-Item	-0.26	2.23	0.51	2.09	0.59

Note. $n_{Male\ E4} = 462$, $n_{Female\ E4} = 89$. $n_{Male\ E5} = 297$, $n_{Female\ E5} = 42$. $d_{F\text{-}M}$ = effect size for Female-Male mean difference. Effect sizes calculated within pay grade as $(M_{Female} - M_{Male})/SD_{Male}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

LeadEx Scores by Race/Ethnicity

Table 3.14 shows mean LeadEx scores by race/ethnicity for Soldiers in each pay grade. Although no significant differences were found between Whites and Hispanics, Whites scored significantly higher than Blacks on both LeadEx composites. Significant Black-White differences were also found in the concurrent validation sample, but the magnitudes of the differences were larger in this sample. Specifically, the effect size statistics for Black-White differences on the 24-item composite in this sample were -0.42 for E4 Soldiers and -0.48 for E5 Soldiers, whereas in the concurrent validation sample the corresponding effects sizes for the 24-item composite were -0.26 and -0.24, respectively.

Although the Black-White differences were larger in this sample, they still fall far below Black-White differences typically associated with traditional tests of cognitive aptitude. For example, effect size statistics for Black-White differences on ASVAB GT (a traditional measure of cognitive aptitude) were -0.75 for E4 Soldiers and -0.78 for E5 Soldiers in this sample. It is also worth noting that the effect size statistics for Hispanic-White differences on the ASVAB GT were -0.65 for E4 Soldiers and -0.55 for E5 Soldiers in this sample. Thus, race differences on LeadEx scores appear far smaller than they are for ASVAB GT, particularly Hispanic-White differences.

Table 3.14. Mean LeadEx Scores by Race/Ethnic Group

			White		Black		Hispanic	
Composite	$d_{B\text{-}W}$	$d_{H\text{-}W}$	*M*	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers								
40-Item	**-0.37**	-0.18	1.94	0.54	1.74	0.69	1.84	0.64
24-Item	**-0.42**	-0.15	1.96	0.57	1.72	0.71	1.88	0.66
E5 Soldiers								
40-Item	**-0.48**	-0.07	2.26	0.42	2.06	0.58	2.23	0.37
24-Item	**-0.48**	-0.06	2.28	0.46	2.06	0.66	2.25	0.45

Note. $n_{White\ E4} = 328$, $n_{Black\ E4} = 110$. $n_{Hispanic\ E4} = 70$, $n_{White\ E5} = 211$, $n_{Black\ E5} = 89$. $n_{Hispanic\ E5} = 28$. $d_{B\text{-}W}$ = effect size for Black-White mean difference. $d_{H\text{-}W}$ = effect size for Hispanic-White mean difference. Effect sizes calculated within pay grade as (mean of non-referent group – M_{White})/SD_{White}. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

LeadEx Scores by MOS

Table 3.15 shows mean LeadEx scores by MOS type for Soldiers in each pay grade. Examination of this table reveals no significant differences between MOS types among E5 Soldiers, and small (yet significant) differences between MOS types among E4 Soldiers. At the E4 pay grade, Soldiers in CA MOS scored significantly lower than Soldiers in other MOS. This finding was similar to results from the concurrent validation effort, where E4 Soldiers in Combat Operations MOS tended to score lower than E4 Soldiers in other MOS.

Table 3.15. Mean LeadEx Scores by MOS Type

Composite	$d_{CS\text{-}CA}$	$d_{CSS\text{-}CA}$	$d_{CSS\text{-}CS}$	Combat Arms		Combat Support		Combat Service Support	
				M	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers									
40-Item	**0.31**	**0.20**	-0.11	1.78	0.61	1.97	0.59	1.90	0.61
24-Item	**0.28**	**0.20**	-0.08	1.80	0.66	1.97	0.61	1.92	0.60
E5 Soldiers									
40-Item	0.08	0.17	0.09	2.15	0.43	2.19	0.53	2.23	0.49
24-Item	0.07	0.07	0.00	2.19	0.48	2.22	0.58	2.23	0.55

Note. $n_{CA\ E4} = 207$, $n_{CS\ E4} = 105$. $n_{CSS\ E4} = 239$, $n_{CA\ E5} = 142$, $n_{CS\ E5} = 44$. $n_{CSS\ E5} = 153$. $d_{CS\text{-}CA}$ = effect size for Combat Support-Combat Arms mean difference. $d_{CSS\text{-}CA}$ = effect size for Combat Service Support-Combat Arms mean difference. $d_{CS\text{-}CA}$ = effect size for Combat Support-Combat Service Support mean difference. Effect sizes calculated within pay grade as (mean of 1st MOS type – mean of 2nd MOS type)/Overall *SD*. Overall *SD* = standard deviation calculated across all Soldiers in the given pay grade (regardless of MOS type). Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Self-Description Inventory (SDI) [12]

The SDI is a 38-item multidimensional forced-choice inventory that measures six temperament constructs: Dependability, Adjustment, Work Orientation, Leadership, Agreeableness, and Physical Conditioning (see Putka, Kilcullen, & White, 2004, for definitions of these constructs). Each item on the SDI presents Soldiers with four statements (a tetrad) that may or may not describe Soldiers' past behavior in common situations. For most of these items, each of the four statements reflects a different construct. Two of these statements are worded positively (often indicating a high standing on each statement's construct of interest) and two are worded negatively (often indicating a low standing on each statement's construct of interest). For each item, Soldiers are asked to select the one statement that is *most like* them and the one statement that is *least like* them. A score for each of the four constructs, represented in a particular item, is generated by assigning a set of points to each statement. Points are assigned based on whether the Soldier's endorsement of that statement (i.e., as most like them or least like them) corresponds to high and low standing on the construct being targeted. SDI scale scores are obtained by computing the mean—across items—of the scores for statements measuring the same construct. Further details on the development and scoring of the SDI can be found in White and Young (1998) and in White (2002).

SDI Scores by Pay Grade

Table 3.16 shows mean SDI scores by pay grade. Like the concurrent validation sample, E5 Soldiers were found to have slightly higher SDI scores than E4 Soldiers. Although most of these differences were statistically significant, the magnitude of the effects was quite small (all < 0.30).

[12] Note that in previous work, the SDI was called the Assessment for Individual Motivation (AIM).

Table 3.16. Mean SDI Scores by Pay Grade

		E4 Soldiers		E5 Soldiers	
Scale	$d_{E5\text{-}E4}$	*M*	*SD*	*M*	*SD*
Dependability	**0.21**	1.17	0.26	1.23	0.22
Adjustment	**0.13**	1.18	0.25	1.21	0.23
Work Orientation	**0.16**	1.23	0.28	1.28	0.24
Leadership	**0.18**	1.25	0.26	1.30	0.24
Agreeableness	**0.19**	1.23	0.25	1.28	0.23
Physical Conditioning	0.07	1.23	0.29	1.25	0.26

Note. n_{E4} = 581, n_{E5} = 346. $d_{E5\text{-}E4}$ = effect size for E5-E4 mean difference. Effect sizes calculated as (M_{E5} - M_{E4})/SD_{E4}. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 3.17 shows correlations among SDI scores by pay grade, as well as internal consistency reliability estimates for each scale. In general, all the SDI scales exhibited adequate levels of internal consistency (potential exceptions were Agreeableness and Dependability for E5 Soldiers). The reliabilities and correlations were similar to those found in the concurrent validation sample.

Table 3.17. SDI Scale Intercorrelations and Reliabilities

Scale	1	2	3	4	5	6
E4 Soldiers						
1. Dependability	(.67)					
2. Adjustment	**.32**	(.75)				
3. Work Orientation	**.45**	**.36**	(.80)			
4. Leadership	**.28**	**.43**	**.63**	(.79)		
5. Agreeableness	**.55**	**.51**	**.43**	**.21**	(.69)	
6. Physical Conditioning	**.29**	**.31**	**.46**	**.12**	**.35**	(.68)
E5 Soldiers						
1. Dependability	(.54)					
2. Adjustment	**.36**	(.73)				
3. Work Orientation	**.29**	**.28**	(.74)			
4. Leadership	**.19**	**.35**	**.63**	(.78)		
5. Agreeableness	**.45**	**.47**	**.35**	**.17**	(.64)	
6. Physical Conditioning	**.20**	**.20**	**.35**	.05	**.27**	(.62)

Note. n_{E4} = 581. n_{E5} = 346. Internal consistency reliability estimates (alpha) are shown in parentheses on the diagonal. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

SDI Scores by Gender

Table 3.18 shows mean SDI scores by gender for Soldiers in each pay grade. No significant gender differences were found for any of the SDI scores among E5 Soldiers. Among E4 Soldiers, two significant effects were found. Like the concurrent validation sample, female E4 Soldiers had significantly higher Dependability scores than male Soldiers. Unlike the concurrent validation sample, female E4 Soldiers had significantly lower Adjustment scores than did male Soldiers. For the most part, these results are similar to those found in the concurrent validation sample.

Table 3.18. Mean SDI Scores by Gender

Scale	d_{F-M}	Male		Female	
		M	*SD*	*M*	*SD*
E4 Soldiers					
Dependability	**0.49**	1.15	0.26	1.28	0.23
Adjustment	**-0.23**	1.19	0.24	1.13	0.25
Work Orientation	0.18	1.23	0.27	1.27	0.29
Leadership	0.01	1.25	0.25	1.25	0.29
Agreeableness	0.15	1.22	0.25	1.26	0.27
Physical Conditioning	-0.02	1.23	0.29	1.23	0.29
E5 Soldiers					
Dependability	0.05	1.22	0.22	1.23	0.23
Adjustment	-0.31	1.22	0.24	1.15	0.20
Work Orientation	-0.27	1.29	0.24	1.22	0.25
Leadership	-0.30	1.31	0.24	1.24	0.24
Agreeableness	-0.10	1.28	0.23	1.26	0.22
Physical Conditioning	-0.18	1.26	0.26	1.21	0.25

Note. $n_{Male\ E4} = 491$, $n_{Female\ E4} = 90$. $n_{Male\ E5} = 303$, $n_{Female\ E5} = 43$. d_{F-M} = effect size for Female-Male mean difference. Effect sizes calculated within pay grade as $(M_{Female} - M_{Male})/SD_{Male}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

SDI Scores by Race/Ethnicity

Table 3.19 shows mean SDI scores by race/ethnicity for Soldiers in each pay grade. Among E4 Soldiers, only one significant race difference was found, and its effect was small (Hispanics scored higher than Whites on Physical Conditioning). Among E5 Soldiers, two significant differences were found. Specifically, Blacks scored significantly lower than Whites on Work Orientation and Leadership. The Black-White differences on these scales (Work Orientation d_{B-W} = -.040; Leadership d_{B-W} = -0.48) were larger than they were in the concurrent validation sample (Work Orientation d_{B-W} = -0.17; Leadership d_{B-W} = -0.13).

Table 3.19. Mean SDI Scores by Race/Ethnic Group

Scale	d_{B-W}	d_{H-W}	White		Black		Hispanic	
			M	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers								
Dependability	0.01	0.06	1.17	0.26	1.17	0.28	1.18	0.26
Adjustment	0.09	0.16	1.17	0.26	1.20	0.21	1.21	0.22
Work Orientation	0.02	0.11	1.23	0.28	1.24	0.26	1.26	0.25
Leadership	0.01	-0.11	1.26	0.26	1.26	0.23	1.23	0.24
Agreeableness	0.04	0.05	1.23	0.26	1.24	0.24	1.24	0.24
Physical Conditioning	0.07	**0.25**	1.22	0.30	1.24	0.25	1.29	0.24
E5 Soldiers								
Dependability	-0.09	0.23	1.23	0.21	1.21	0.22	1.28	0.24
Adjustment	-0.22	-0.06	1.23	0.24	1.18	0.21	1.22	0.25
Work Orientation	**-0.40**	0.13	1.30	0.23	1.21	0.25	1.33	0.29
Leadership	**-0.48**	-0.16	1.33	0.23	1.22	0.24	1.29	0.27
Agreeableness	-0.09	-0.03	1.29	0.24	1.26	0.22	1.28	0.20
Physical Conditioning	0.01	-0.23	1.25	0.27	1.26	0.24	1.19	0.30

Note. $n_{White\ E4} = 342$, $n_{Black\ E4} = 118$. $n_{Hispanic\ E4} = 73$, $n_{White\ E5} = 215$, $n_{Black\ E5} = 92$. $n_{Hispanic\ E5} = 28$. d_{B-W} = effect size for Black-White mean difference. d_{H-W} = effect size for Hispanic-White mean difference. Effect sizes calculated within pay grade as (mean of non-referent group – M_{White})/SD_{White}. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

SDI Scores by MOS

Table 3.20 shows mean SDI scores by MOS type for Soldiers in each pay grade. Examination of this table reveals few significant MOS differences. At the E4 pay grade, Soldiers in CSS MOS had significantly higher Adjustment scores than Soldiers in other MOS, as well as significantly higher Dependability scores than Soldiers in CA MOS. Although these effects were significant, they were small (all < 0.30). At the E5 pay grade, Soldiers in CA MOS had significantly higher Adjustment and Leadership scores than Soldiers in CSS MOS and significantly higher Work Orientation scores than Soldiers in CS MOS. Like differences found at the E4 level, these effects tended to be small. These results are similar to those found in the concurrent validation sample.

Table 3.20. Mean SDI Scores by MOS Type

				Combat Arms		Combat Support		Combat Service Support	
Scale	$d_{CS\text{-}CA}$	$d_{CSS\text{-}CA}$	$d_{CSS\text{-}CS}$	*M*	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers									
Dependability	0.07	**0.28**	0.21	1.13	0.26	1.15	0.28	1.21	0.25
Adjustment	-0.05	0.01	0.05	1.18	0.24	1.17	0.27	1.18	0.24
Work Orientation	-0.05	**0.24**	**0.29**	1.21	0.28	1.19	0.27	1.27	0.27
Leadership	-0.09	0.02	0.10	1.25	0.26	1.23	0.29	1.26	0.24
Agreeableness	0.05	0.09	0.03	1.22	0.25	1.23	0.23	1.24	0.26
Physical Conditioning	-0.09	0.03	0.12	1.23	0.28	1.21	0.31	1.24	0.28
E5 Soldiers									
Dependability	0.11	0.21	0.10	1.20	0.22	1.23	0.24	1.25	0.20
Adjustment	-0.03	**-0.24**	-0.22	1.24	0.23	1.23	0.21	1.18	0.23
Work Orientation	**-0.37**	-0.13	0.24	1.30	0.23	1.21	0.28	1.27	0.24
Leadership	-0.27	**-0.35**	-0.07	1.35	0.23	1.28	0.25	1.26	0.24
Agreeableness	0.04	0.03	-0.01	1.27	0.24	1.28	0.25	1.28	0.22
Physical Conditioning	-0.26	-0.02	0.24	1.26	0.26	1.19	0.24	1.26	0.27

Note. $n_{CA\ E4} = 222$, $n_{CS\ E4} = 108$. $n_{CSS\ E4} = 251$, $n_{CA\ E5} = 144$, $n_{CS\ E5} = 45$. $n_{CSS\ E5} = 157$. $d_{CS\text{-}CA}$ = effect size for Combat Support-Combat Arms mean difference. $d_{CSS\text{-}CA}$ = effect size for Combat Service Support-Combat Arms mean difference. $d_{CS\text{-}CA}$ = effect size for Combat Support-Combat Service Support mean difference. Effect sizes calculated within pay grade as (mean of 1st MOS type – mean of 2nd MOS type)/Overall *SD*. Overall *SD* = standard deviation calculated across all Soldiers in the given pay grade (regardless of MOS type). Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Information Questionnaire-II (IQ-II)[13]

The concurrent validation version of the IQ-II is a 156-item measure of eight temperament constructs: Hostility to Authority, Manipulativeness, Social Maturity, Tolerance for Ambiguity, Openness, Emergent Leadership, Social Perceptiveness, and Interpersonal Skill (see Putka et al., 2004, for definitions of these constructs). The longitudinal research version of this instrument did not include two of these scales (i.e., Openness and Social Maturity) due to modest results in the concurrent validation effort (Knapp, McCloy, & Heffner, 2004).[14] The items that

[13] In previous project reports, the IQ-II was called the Biographical Information Questionnaire (BIQ).

[14] Additional scales were included in the longitudinal version of the instrument, but are not discussed here because they were not part of the concurrent validation effort. However, both versions contain 156 items.

constitute the IQ-II reflect prior behaviors and reactions to specific life events indicative of the targeted psychological constructs. IQ-II items were drawn from existing biodata instruments the Army has used for operational and research purposes (see Putka et al., 2004, for a review). Soldiers complete the IQ-II by indicating the extent to which each item describes their past behavior using a five-option Likert rating scale. Response options on the IQ-II were scored rationally, based on the hypothesized relation of the item responses to the underlying psychological construct. Scores for each IQ-II scale were calculated by calculating the mean of the Soldiers' responses across items corresponding to each construct.

IQ-II Scores by Pay Grade

Table 3.21 shows mean IQ-II scores by pay grade. As in the concurrent validation sample, E5 Soldiers were found to have significantly higher scores on Interpersonal Skills and Emergent Leadership and significantly lower scores on Manipulativeness and Hostility to Authority compared to E4 Soldiers. Though these effects were significant, they were all small (the largest effect size was 0.35).

Table 3.21. Mean IQ-II Scores by Pay Grade

		E4 Soldiers		E5 Soldiers	
Scale	$d_{E5\text{-}E4}$	*M*	*SD*	*M*	*SD*
Tolerance for Ambiguity	0.01	3.15	0.42	3.16	0.42
Interpersonal Skills	**0.22**	3.14	0.45	3.24	0.44
Social Perceptiveness	-0.08	3.54	0.53	3.50	0.49
Emergent Leadership	**0.32**	3.33	0.57	3.51	0.48
Manipulativeness	**-0.35**	2.35	0.49	2.18	0.47
Hostility to Authority	**-0.30**	3.11	0.57	2.93	0.52

Note. $n_{E4} = 579$, $n_{E5} = 345$. $d_{E5\text{-}E4}$ = effect size for E5-E4 mean difference. Effect sizes calculated as $(M_{E5} - M_{E4})/SD_{E4}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 3.22 shows correlations among IQ-II scores by pay grade, as well as internal consistency reliability estimates for each scale. As in the concurrent validation sample, the Tolerance for Ambiguity and Interpersonal Skills scales exhibited lesser levels of reliability. With the potential exception of the Hostility to Authority scale among E5 Soldiers, all other IQ-II scales showed relatively strong levels of reliability. In general, the reliabilities and correlations among IQ-II scales were similar to those found in the concurrent validation sample.

IQ-II Scores by Gender

Table 3.23 shows mean IQ-II scores by gender for Soldiers in each pay grade. Few significant gender differences were found on the IQ-II scales. As in the concurrent validation sample, female E4 Soldiers scored significantly lower than male E4 Soldiers on Hostility to Authority. Unlike the concurrent validation sample, female E4 Soldiers scored significantly lower than male E4 Soldiers on Emergent Leadership, and female E5 Soldiers scored significantly lower than male E5 Soldiers on Tolerance for Ambiguity. The prevalence of non-significant gender differences on the IQ-II resembled findings from the concurrent validation effort.

Table 3.22. IQ-II Scale Intercorrelations and Reliabilities

Scale	1	2	3	4	5	6
E4 Soldiers						
1. Tolerance for Ambiguity	(.52)					
2. Interpersonal Skills	**.33**	(.57)				
3. Social Perceptiveness	**.37**	**.24**	(.85)			
4. Emergent Leadership	**.46**	**.35**	**.66**	(.84)		
5. Manipulativeness	**-.34**	**-.44**	**-.15**	**-.28**	(.73)	
6. Hostility to Authority	**-.20**	**-.40**	**.15**	.01	**.51**	(.74)
E5 Soldiers						
1. Tolerance for Ambiguity	(.54)					
2. Interpersonal Skills	**.32**	(.54)				
3. Social Perceptiveness	**.32**	**.32**	(.82)			
4. Emergent Leadership	**.44**	**.33**	**.60**	(.79)		
5. Manipulativeness	**-.32**	**-.47**	**-.22**	**-.29**	(.75)	
6. Hostility to Authority	**-.21**	**-.46**	.02	-.07	**.54**	(.66)

Note. n_{E4} = 579. n_{E5} = 345. Internal consistency reliability estimates (alpha) are shown in parentheses on the diagonal. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

Table 3.23. Mean IQ-II Scores by Gender

		Male		Female	
Scale	d_{F-M}	*M*	*SD*	*M*	*SD*
E4 Soldiers					
Tolerance for Ambiguity	-0.14	3.16	0.42	3.10	0.42
Interpersonal Skills	-0.13	3.15	0.45	3.09	0.47
Social Perceptiveness	-0.07	3.55	0.53	3.51	0.54
Emergent Leadership	**-0.24**	3.35	0.56	3.22	0.59
Manipulativeness	-0.20	2.37	0.50	2.27	0.46
Hostility to Authority	**-0.42**	3.14	0.56	2.91	0.56
E5 Soldiers					
Tolerance for Ambiguity	**-0.43**	3.18	0.41	3.00	0.48
Interpersonal Skills	-0.21	3.25	0.44	3.15	0.42
Social Perceptiveness	0.11	3.50	0.47	3.55	0.63
Emergent Leadership	-0.25	3.53	0.47	3.41	0.54
Manipulativeness	0.06	2.18	0.46	2.20	0.54
Hostility to Authority	0.07	2.93	0.51	2.97	0.59

Note. $n_{Male\ E4}$ = 486, $n_{Female\ E4}$ = 93. $n_{Male\ E5}$ = 302, $n_{Female\ E5}$ = 43. d_{F-M} = effect size for Female-Male mean difference. Effect sizes calculated within pay grade as $(M_{Female} - M_{Male})/SD_{Male}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

IQ-II Scores by Race/Ethnicity

Table 3.24 shows mean IQ-II scores by race/ethnicity for Soldiers in each pay grade. Among E4 Soldiers, two small but significant race differences emerged, and both were on the Social Perceptiveness scale. Specifically, Whites' Social Perceptiveness scores were significantly lower than Blacks' scores but significantly higher than Hispanics' scores. Among

E5 Soldiers, no significant differences were found between Whites and Hispanics, though three significant differences were found between Whites and Blacks. As in the concurrent validation sample, we found that White Soldiers had significantly higher Tolerance for Ambiguity scores than Black Soldiers. Unlike the concurrent validation sample, we also found that White Soldiers had significantly higher scores than Black Soldiers on Interpersonal Skills and Emergent Leadership; however, these differences are relatively small.

Table 3.24. Mean IQ-II Scores by Race/Ethnic Group

			White		Black		Hispanic	
Scale	$d_{B\text{-}W}$	$d_{H\text{-}W}$	*M*	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers								
Tolerance for Ambiguity	-0.19	-0.19	3.19	0.43	3.11	0.39	3.11	0.42
Interpersonal Skills	-0.16	-0.14	3.16	0.47	3.08	0.44	3.09	0.46
Social Perceptiveness	**0.25**	**-0.25**	3.54	0.55	3.67	0.45	3.40	0.51
Emergent Leadership	0.10	-0.19	3.34	0.54	3.39	0.57	3.23	0.58
Manipulativeness	0.05	0.02	2.34	0.47	2.37	0.58	2.35	0.47
Hostility to Authority	-0.02	-0.25	3.12	0.55	3.11	0.62	2.99	0.56
E5 Soldiers								
Tolerance for Ambiguity	**-0.54**	-0.26	3.23	0.41	3.01	0.41	3.12	0.41
Interpersonal Skills	**-0.36**	-0.19	3.29	0.45	3.13	0.41	3.21	0.40
Social Perceptiveness	-0.13	0.10	3.52	0.50	3.45	0.48	3.56	0.42
Emergent Leadership	**-0.29**	0.01	3.55	0.47	3.41	0.49	3.55	0.54
Manipulativeness	0.21	0.08	2.15	0.45	2.24	0.52	2.18	0.45
Hostility to Authority	0.14	-0.19	2.92	0.52	2.99	0.52	2.82	0.50

Note. $n_{White\ E4} = 338$, $n_{Black\ E4} = 120$. $n_{Hispanic\ E4} = 72$, $n_{White\ E5} = 213$, $n_{Black\ E5} = 93$. $n_{Hispanic\ E5} = 28$. $d_{B\text{-}W}$ = effect size for Black-White mean difference. $d_{H\text{-}W}$ = effect size for Hispanic-White mean difference. Effect sizes calculated within pay grade as (mean of non-referent group – M_{White})/SD_{White}. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

IQ-II Scores by MOS

Table 3.25 shows mean IQ-II scores by MOS type for Soldiers in each pay grade. Examination of this table reveals few significant MOS differences on the IQ-II scales. At the E4 pay grade, Soldiers in CS MOS had significantly higher Tolerance for Ambiguity and Interpersonal Skills scores than Soldiers in CSS MOS. Unlike this sample, concurrent validation sample E4 Soldiers in a CMF dominated by CA MOS were significantly higher than most other CMFs on Manipulativeness and Hostility to Authority. At the E5 pay grade in the longitudinal validation sample, Soldiers in CA MOS had significantly higher Tolerance for Ambiguity scores than Soldiers in CSS MOS. Although these effects were significant, they were quite small (all < 0.30). Overall, these results are similar to those found in the concurrent sample.

Table 3.25. Mean IQ-II Scores by MOS Type

Scale	$d_{CS\text{-}CA}$	$d_{CSS\text{-}CA}$	$d_{CSS\text{-}CS}$	Combat Arms		Combat Support		Combat Service Support	
				M	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers									
Tolerance for Ambiguity	0.17	-0.10	**-0.27**	3.16	0.42	3.23	0.49	3.12	0.39
Interpersonal Skills	0.21	-0.05	**-0.26**	3.13	0.44	3.23	0.45	3.11	0.47
Social Perceptiveness	0.15	-0.05	-0.21	3.54	0.55	3.62	0.53	3.51	0.51
Emergent Leadership	-0.03	0.00	0.03	3.34	0.57	3.32	0.57	3.34	0.57
Manipulativeness	-0.08	-0.16	-0.08	2.40	0.48	2.35	0.52	2.32	0.49
Hostility to Authority	-0.13	-0.05	0.08	3.13	0.59	3.06	0.55	3.10	0.56
E5 Soldiers									
Tolerance for Ambiguity	-0.18	**-0.27**	-0.09	3.22	0.42	3.14	0.41	3.10	0.42
Interpersonal Skills	0.29	0.09	-0.20	3.20	0.45	3.33	0.45	3.24	0.41
Social Perceptiveness	-0.04	0.00	0.05	3.50	0.45	3.48	0.48	3.51	0.52
Emergent Leadership	-0.14	-0.17	-0.03	3.56	0.47	3.49	0.51	3.48	0.49
Manipulativeness	0.18	0.07	-0.11	2.15	0.46	2.24	0.51	2.19	0.47
Hostility to Authority	-0.08	0.03	0.11	2.93	0.53	2.89	0.56	2.95	0.50

Note. $n_{CA\ E4} = 218$, $n_{CS\ E4} = 108$. $n_{CSS\ E4} = 253$, $n_{CA\ E5} = 143$, $n_{CS\ E5} = 44$. $n_{CSS\ E5} = 158$. $d_{CS\text{-}CA}$ = effect size for Combat Support-Combat Arms mean difference. $d_{CSS\text{-}CA}$ = effect size for Combat Service Support-Combat Arms mean difference. $d_{CS\text{-}CA}$ = effect size for Combat Support-Combat Service Support mean difference. Effect sizes calculated within pay grade as (mean of 1st MOS type – mean of 2nd MOS type)/Overall *SD*. Overall *SD* = standard deviation calculated across all Soldiers in the given pay grade (regardless of MOS type). Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Work Suitability Inventory (WSI)[15]

The Work Suitability Inventory (WSI) comprises 16 statements that describe temperament-related work requirements. All but one of the statements are based on the Work Styles portion of the O*NET content model (Borman, Kubisiak, & Schneider, 1999).[16] The WSI presents Soldiers with a computerized card-sorting task. Sixteen cards are displayed on the screen, and each card contains one of the work characteristic statements. Soldiers must sort the 16 cards in terms of how well they think they would perform the type of work described by the cards. Cards containing types of work that they think they would perform well are ranked highest; cards containing types of work that they think they would perform worst are ranked lowest. Respondents sort the 16 cards by using the computer mouse to drag and drop the cards into 16 "ranking" boxes outlined on the screen.

Scoring of the WSI

The score assigned to each WSI trait (each trait is represented by a single card) is computed as 17 minus its rank (which can range from 1 to 16). This scoring method results in complete ipsativity (i.e., the sum of each Soldier's WSI trait scores is the same for all Soldiers). From a

[15] The WSI was not administered during the NCO21 concurrent validation effort. It was developed as part of another ARI project (Select21).

[16] The statement that was not taken from the O*NET addresses cultural tolerance.

statistical perspective, such ipsativity is undesirable; nevertheless, it provides a means for describing Soldiers' rank ordering of all 16 traits assessed on the WSI. Therefore, we adopted this scoring option for all WSI analyses presented in this report. However, it is important to note that under this scoring option, the magnitude of WSI scores do not necessarily reflect Soldiers' standing on a given personality trait, but rather how well they feel they could perform work that requires a certain personality trait relative to other types of work. For a discussion of other scoring options for this instrument see Knapp, Sager, & Tremble (2005).

WSI Scores by Pay Grade

Table 3.26 shows mean WSI scores by pay grade. Although five statistically significant pay grade differences were found across the 16 WSI scales, all of these effects were small (< 0.30). Compared to E5 Soldiers, E4 Soldiers viewed themselves as more capable of performing work requiring Innovation, Persistence, and Cultural Tolerance and less capable of performing work requiring Leadership Orientation and Stress Tolerance. In general, Soldiers at both pay grades tended to view themselves as being most capable of performing work requiring Achievement/Effort, Attention to Detail, and Leadership Orientation, and least capable of performing work that required Persistence and Stress Tolerance.

Table 3.26 Mean WSI Scores by Pay Grade

		E4 Soldiers		E5 Soldiers	
Scale	$d_{E5\text{-}E4}$	*M*	*SD*	*M*	*SD*
Achievement/Effort	0.06	10.68	4.31	10.95	4.35
Adaptability/Flexibility	0.10	9.80	4.22	10.20	4.19
Attention to Detail	0.06	10.18	4.33	10.45	3.97
Concern for Others	0.07	7.48	4.60	7.82	4.76
Cooperation	-0.13	7.56	4.20	7.03	4.44
Dependability	-0.02	8.91	4.10	8.84	4.07
Energy	-0.12	8.81	4.31	8.27	4.27
Independence	0.03	9.68	5.22	9.83	4.96
Initiative	0.08	7.90	4.19	8.25	4.25
Innovation	**-0.20**	9.96	4.22	9.11	4.16
Leadership Orientation	**0.25**	10.08	4.51	11.22	4.05
Persistence	**-0.19**	6.56	3.93	5.81	3.89
Self-Control	-0.03	7.06	4.18	6.94	4.23
Social Orientation	-0.04	7.89	4.61	7.70	4.28
Stress Tolerance	**0.20**	5.11	4.01	5.92	4.34
Cultural Tolerance	**-0.15**	8.35	4.75	7.65	4.70

Note. n_{E4} = 540, n_{E5} = 322. $d_{E5\text{-}E4}$ = effect size for E5-E4 mean difference. Effect sizes calculated as (M_{E5} - M_{E4})/SD_{E4}. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 3.27 shows correlations among WSI scores by pay grade. Perhaps the most striking feature of this correlation matrix is the preponderance of negative correlations. Although negative correlations between such traits are unusual, the ipsative nature of the WSI scores renders them expected results (Hicks, 1970). For example, Soldiers who indicated they were more capable of performing work requiring Achievement/Effort and Attention to Detail (relative to other types of work) tended to indicate they were less capable of performing work requiring

Social Orientation and Cultural Tolerance (which would suggest a task-oriented vs. person-oriented interpretation of the data). Interpreted through this lens, many of the negative correlations observed in this table make conceptual sense. Further, the positive correlations in the matrix were generally found for traits that one would expect to be most positively correlated when assessed using a non-ipsative measure (e.g., Concern for Others with Cooperation; Achievement/Effort with Dependability and Attention to Detail).

Table 3.27. WSI Scale Intercorrelations

Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
E4 Soldiers															
1. Achievement/Effort															
2. Adaptability/Flexibility	**.10**														
3. Attention to Detail	**.14**	.05													
4. Concern for Others	-.06	**.08**	.06												
5. Cooperation	**-.08**	.07	.00	**.30**											
6. Dependability	**.08**	-.04	**.16**	**-.08**	.03										
7. Energy	.01	**-.17**	-.06	**-.11**	-.07	.06									
8. Independence	**-.11**	-.06	-.07	**-.15**	**-.13**	**-.09**	-.05								
9. Initiative	-.02	**-.10**	**-.12**	**-.15**	**-.17**	-.03	**-.11**	-.01							
10. Innovation	**-.13**	**-.08**	**-.26**	**-.17**	**-.20**	**-.23**	-.04	-.02	.04						
11. Leadership Orientation	**-.11**	**-.19**	**-.13**	**-.25**	**-.19**	-.02	-.01	**-.09**	-.01	**.10**					
12. Persistence	**-.11**	**-.16**	-.07	**-.19**	**-.12**	**-.12**	**-.10**	-.02	.06	.05	-.04				
13. Self-Control	**-.18**	**-.21**	**-.10**	**-.09**	**-.13**	**-.19**	-.02	**-.12**	**-.11**	-.04	.01	.04			
14. Social Orientation	**-.18**	**-.13**	**-.25**	.01	-.05	**-.16**	**-.13**	**-.21**	-.02	-.01	.03	-.04	.03		
15. Stress Tolerance	**-.12**	**-.12**	**-.12**	**-.25**	**-.24**	-.07	-.06	.01	-.02	-.02	.04	.02	**.17**	**-.09**	
16. Cultural Tolerance	**-.16**	-.03	**-.19**	.00	-.02	**-.22**	**-.14**	-.07	**-.17**	.04	**-.14**	**-.10**	.00	**.13**	-.05
E5 Soldiers															
1. Achievement/Effort															
2. Adaptability/Flexibility	**.12**														
3. Attention to Detail	**.23**	**.16**													
4. Concern for Others	-.02	.01	-.06												
5. Cooperation	.06	.08	-.08	**.31**											
6. Dependability	.04	**-.15**	.08	-.06	-.04										
7. Energy	-.09	**-.11**	-.01	-.07	-.05	**.10**									
8. Independence	-.06	**-.19**	-.07	**-.17**	**-.16**	.03	.01								
9. Initiative	-.08	**-.14**	-.08	**-.17**	**-.14**	.00	-.03	-.06							
10. Innovation	**-.13**	.00	-.05	**-.23**	**-.18**	**-.18**	**-.15**	.05	**.09**						
11. Leadership Orientation	**-.14**	**-.27**	-.09	**-.24**	**-.21**	-.06	.02	-.01	.03	.00					
12. Persistence	-.08	**-.10**	.03	**-.14**	**-.20**	.00	-.07	.06	-.04	**.10**	**-.14**				
13. Self-Control	**-.20**	**-.12**	**-.24**	**-.12**	**-.17**	**-.11**	**-.12**	**-.13**	-.02	**-.11**	**.10**	-.07			
14. Social Orientation	**-.25**	**-.12**	**-.15**	.00	-.05	**-.14**	**-.16**	**-.20**	**-.14**	-.07	**.13**	**-.16**	.07		
15. Stress Tolerance	**-.19**	**-.12**	**-.24**	**-.26**	**-.22**	**-.19**	-.09	-.02	-.03	-.05	.02	.08	**.22**	.03	
16. Cultural Tolerance	**-.20**	-.01	**-.30**	.09	-.02	**-.25**	**-.18**	**-.18**	**-.14**	-.03	-.07	**-.14**	.04	**.19**	.07

Note. $n_{E4} = 540$. $n_{E5} = 322$. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

WSI Scores by Gender

Table 3.28 shows mean WSI scores by gender for Soldiers in each pay grade. At the E4 pay grade, female Soldiers viewed themselves as significantly more capable of performing work requiring Attention to Detail and Cultural Tolerance (relative to other types of work), and significantly less capable of performing work requiring Persistence and Stress Tolerance (again, relative to other types of work) compared to male Soldiers. At the E5 pay grade, female Soldiers viewed themselves as significantly more capable of performing work requiring Concern for Others, and significantly less capable of performing work requiring Independence and Leadership Orientation compared to male Soldiers.

Table 3.28. Mean WSI Scores by Gender

		Male		Female	
Scale	$d_{F\text{-}M}$	*M*	*SD*	*M*	*SD*
E4 Soldiers					
Achievement/Effort	-0.02	10.69	4.36	10.60	4.04
Adaptability/Flexibility	0.07	9.76	4.24	10.04	4.14
Attention to Detail	**0.31**	9.97	4.36	11.32	4.01
Concern for Others	0.23	7.31	4.56	8.35	4.75
Cooperation	-0.03	7.58	4.23	7.46	4.06
Dependability	-0.10	8.98	4.10	8.56	4.12
Energy	-0.13	8.90	4.38	8.32	3.89
Independence	0.08	9.62	5.19	10.01	5.39
Initiative	-0.02	7.91	4.22	7.85	4.06
Innovation	-0.12	10.04	4.22	9.52	4.20
Leadership Orientation	-0.21	10.22	4.45	9.29	4.77
Persistence	**-0.26**	6.72	4.01	5.68	3.37
Self-Control	-0.14	7.16	4.22	6.55	3.92
Social Orientation	0.16	7.77	4.55	8.48	4.90
Stress Tolerance	**-0.36**	5.34	4.09	3.87	3.33
Cultural Tolerance	**0.44**	8.02	4.76	10.09	4.35
E5 Soldiers					
Achievement/Effort	0.30	10.79	4.39	12.11	3.90
Adaptability/Flexibility	0.20	10.10	4.24	10.95	3.77
Attention to Detail	-0.05	10.48	3.97	10.29	4.00
Concern for Others	**0.46**	7.56	4.71	9.74	4.79
Cooperation	0.27	6.89	4.45	8.08	4.33
Dependability	0.11	8.78	4.13	9.24	3.63
Energy	-0.22	8.38	4.24	7.45	4.49
Independence	**-0.36**	10.04	4.92	8.26	5.04
Initiative	0.10	8.20	4.22	8.63	4.47
Innovation	-0.27	9.24	4.18	8.13	3.93
Leadership Orientation	**-0.37**	11.39	3.94	9.95	4.66
Persistence	-0.26	5.93	3.91	4.92	3.66
Self-Control	-0.05	6.96	4.26	6.74	4.00
Social Orientation	0.09	7.65	4.21	8.03	4.79
Stress Tolerance	-0.28	6.07	4.41	4.82	3.63
Cultural Tolerance	0.25	7.51	4.71	8.68	4.63

Note. $n_{\text{Male E4}} = 455$, $n_{\text{Female E4}} = 85$. $n_{\text{Male E5}} = 284$, $n_{\text{Female E5}} = 38$. $d_{F\text{-}M}$ = effect size for Female-Male mean difference. Effect sizes calculated within pay grade as $(M_{\text{Female}} - M_{\text{Male}})/SD_{\text{Male}}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

WSI Scores by Race/Ethnicity

Table 3.29 shows mean WSI scores by race/ethnicity for Soldiers in each pay grade. At the E4 and E5 pay grades, several significant race differences were found. At the E4 level, White Soldiers viewed themselves as significantly more capable of performing work requiring Independence, Initiative, and Stress Tolerance (relative to other types of work) compared to

Table 3.29. Mean WSI Scores by Race/Ethnic Group

			White		Black		Hispanic	
Scale	$d_{B\text{-}W}$	$d_{H\text{-}W}$	*M*	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers								
Achievement/Effort	-0.05	0.00	10.67	4.37	10.46	4.18	10.65	4.16
Adaptability/Flexibility	0.11	0.16	9.52	4.25	9.97	3.88	10.19	4.29
Attention to Detail	0.11	0.09	10.02	4.39	10.51	4.20	10.40	4.34
Concern for Others	0.22	0.17	7.06	4.70	8.07	4.52	7.85	4.58
Cooperation	0.18	**0.30**	7.11	4.18	7.88	4.05	8.36	4.31
Dependability	0.05	0.13	8.67	4.22	8.89	3.73	9.22	4.28
Energy	-0.05	0.04	8.82	4.46	8.60	4.19	9.01	3.99
Independence	**-0.26**	**-0.46**	10.41	5.17	9.06	5.16	8.06	5.10
Initiative	**-0.26**	**-0.26**	8.33	4.19	7.24	4.27	7.25	3.89
Innovation	-0.09	-0.10	10.18	4.22	9.81	4.22	9.75	3.92
Leadership Orientation	0.15	0.12	9.87	4.45	10.54	4.51	10.39	4.75
Persistence	**-0.29**	-0.07	6.89	4.10	5.71	3.58	6.61	3.66
Self-Control	-0.04	-0.13	7.19	4.21	7.04	4.28	6.64	4.02
Social Orientation	0.20	0.14	7.65	4.59	8.58	4.63	8.28	4.59
Stress Tolerance	**-0.34**	**-0.34**	5.72	4.03	4.35	4.11	4.36	4.01
Cultural Tolerance	**0.31**	0.24	7.89	4.46	9.28	5.05	8.97	5.14
E5 Soldiers								
Achievement/Effort	0.03	-0.01	10.86	4.36	10.98	4.45	10.82	4.32
Adaptability/Flexibility	0.03	0.09	10.15	4.17	10.26	4.33	10.54	4.17
Attention to Detail	**-0.30**	-0.38	10.89	3.90	9.74	4.06	9.43	3.77
Concern for Others	**0.43**	0.19	7.13	4.64	9.14	4.77	8.00	4.86
Cooperation	**0.50**	0.13	6.37	4.25	8.48	4.53	6.93	4.90
Dependability	0.01	-0.21	8.89	3.99	8.92	4.43	8.07	3.80
Energy	0.06	0.35	8.05	4.38	8.30	4.11	9.57	3.98
Independence	**-0.29**	-0.07	10.31	4.84	8.91	5.08	9.96	4.85
Initiative	-0.25	**-0.56**	8.82	4.04	7.82	4.08	6.57	5.28
Innovation	-0.19	0.02	9.39	4.44	8.53	3.56	9.50	3.75
Leadership Orientation	0.00	0.15	11.10	4.08	11.11	4.16	11.71	3.75
Persistence	-0.13	-0.05	6.01	3.80	5.51	4.21	5.82	3.75
Self-Control	-0.19	-0.29	7.26	4.45	6.41	3.89	5.96	3.42
Social Orientation	0.16	-0.21	7.53	4.21	8.21	4.43	6.64	4.26
Stress Tolerance	**-0.28**	0.13	6.29	4.37	5.07	4.20	6.86	4.32
Cultural Tolerance	**0.37**	**0.59**	6.94	4.51	8.62	4.69	9.61	5.28

Note. $n_{White\ E4} = 315$ $n_{Black\ E4} = 108$. $n_{Hispanic\ E4} = 72$, $n_{White\ E5} = 199$. $n_{Black\ E5} = 87$. $n_{Hispanic\ E5} = 28$. $d_{B\text{-}W}$ = effect size for Black-White mean difference. $d_{H\text{-}W}$ = effect size for Hispanic-White mean difference. Effect sizes calculated within pay grade as (mean of non-referent group – M_{White})/SD_{White}. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Black and Hispanic Soldiers, and significantly more capable of performing work requiring Persistence (again, relative to other types of work) compared to Black Soldiers. Conversely, White Soldiers viewed themselves as significantly less capable of performing work requiring Cooperation (relative to other types of work) compared to Hispanic Soldiers, and significantly less capable of performing work requiring Cultural Tolerance compared to Black Soldiers.

At the E5 level, White Soldiers viewed themselves as significantly less capable of performing work requiring Cultural Tolerance (relative to other types of work) compared to Black and Hispanic Soldiers, and significantly less capable of performing work requiring Cooperation and Concern for Others compared to Black Soldiers. Conversely, White Soldiers viewed themselves as significantly more capable of performing work requiring Attention to Detail, Independence, and Stress Tolerance (again, relative to other types of work) compared to Black Soldiers, and significantly more capable of performing work requiring Initiative compared to Hispanic Soldiers.

WSI Scores by MOS

Table 3.30 shows mean WSI scores by MOS type for Soldiers in each pay grade. Examination of this table reveals few MOS differences on the WSI scales. At the E4 pay grade, Soldiers in CSS MOS viewed themselves as significantly more capable of performing work requiring Persistence (relative to other types of work) compared to Soldiers in CS MOS, and significantly more capable of performing work requiring Cultural Tolerance (again, relative to other types of work) compared to Soldiers in CA MOS. Conversely, Soldiers in CSS MOS viewed themselves as significantly less capable of performing work requiring Stress Tolerance (relative to other types of work) compared to Soldiers in CA MOS. Further, Soldiers in CS MOS viewed themselves as significantly less capable of performing work requiring Leadership Orientation compared to Soldiers in CA MOS.

Table 3.30. Mean WSI Scores by MOS Type

				Combat Arms		Combat Support		Combat Service Support	
Scale	$d_{CS\text{-}CA}$	$d_{CSS\text{-}CA}$	$d_{CSS\text{-}CS}$	*M*	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers									
Achievement/Effort	-0.06	-0.13	-0.08	10.97	4.24	10.73	4.36	10.40	4.35
Adaptability/Flexibility	0.21	0.03	-0.18	9.58	4.21	10.45	4.18	9.70	4.24
Attention to Detail	0.01	0.18	0.17	9.85	4.30	9.88	4.53	10.61	4.23
Concern for Others	0.06	-0.01	-0.07	7.44	4.55	7.71	4.71	7.41	4.61
Cooperation	0.13	-0.06	-0.19	7.57	4.13	8.10	4.30	7.31	4.20
Dependability	-0.20	-0.09	0.11	9.23	4.13	8.41	3.82	8.86	4.18
Energy	0.03	-0.02	-0.06	8.82	4.45	8.96	4.19	8.72	4.24
Independence	0.06	0.01	-0.05	9.60	5.29	9.90	5.02	9.65	5.25
Initiative	0.01	-0.08	-0.09	8.03	4.37	8.09	4.10	7.71	4.07
Innovation	-0.02	-0.17	-0.15	10.27	4.33	10.20	4.32	9.57	4.06
Leadership Orientation	**-0.28**	-0.08	0.20	10.49	4.20	9.21	4.96	10.11	4.52
Persistence	-0.09	0.19	**0.27**	6.31	3.81	5.97	3.60	7.05	4.13
Self-Control	-0.06	0.03	0.09	7.05	4.15	6.81	4.18	7.19	4.22
Social Orientation	0.15	0.15	0.00	7.46	4.62	8.15	4.72	8.14	4.54
Stress Tolerance	-0.10	**-0.29**	-0.19	5.68	4.29	5.28	3.86	4.52	3.75
Cultural Tolerance	0.10	**0.29**	0.19	7.65	4.54	8.14	4.81	9.05	4.83

Table 3.30. (Continued)

Scale	$d_{CS\text{-}CA}$	$d_{CSS\text{-}CA}$	$d_{CSS\text{-}CS}$	Combat Arms *M*	Combat Arms *SD*	Combat Support *M*	Combat Support *SD*	Combat Service Support *M*	Combat Service Support *SD*
E5 Soldiers									
Achievement/Effort	-0.15	0.06	0.20	10.93	4.45	10.30	4.35	11.17	4.26
Adaptability/Flexibility	-0.02	-0.16	-0.13	10.51	4.08	10.41	4.21	9.85	4.28
Attention to Detail	0.08	0.21	0.13	10.04	4.03	10.36	3.70	10.87	3.96
Concern for Others	0.19	0.21	0.02	7.26	4.58	8.16	5.00	8.25	4.84
Cooperation	0.00	0.20	0.20	6.65	4.51	6.64	4.20	7.52	4.43
Dependability	0.06	0.03	-0.03	8.75	4.00	9.00	3.95	8.87	4.20
Energy	0.03	-0.03	-0.06	8.32	4.24	8.43	4.08	8.18	4.38
Independence	0.14	-0.07	-0.21	9.90	5.00	10.59	4.99	9.53	4.92
Initiative	-0.04	-0.09	-0.05	8.45	4.07	8.27	4.65	8.06	4.31
Innovation	0.10	-0.12	-0.23	9.28	4.11	9.70	4.68	8.77	4.04
Leadership Orientation	**-0.53**	**-0.30**	0.23	12.04	3.80	9.91	4.16	10.85	4.12
Persistence	0.24	0.05	-0.19	5.60	3.72	6.55	3.80	5.80	4.07
Self-Control	-0.03	0.13	0.16	6.71	4.37	6.59	4.23	7.26	4.09
Social Orientation	-0.11	0.04	0.15	7.68	4.24	7.20	4.42	7.86	4.29
Stress Tolerance	**-0.37**	**-0.39**	-0.03	6.89	4.42	5.30	4.39	5.18	4.09
Cultural Tolerance	0.34	0.21	-0.13	7.00	4.59	8.59	4.83	7.99	4.72

Note. $n_{CA\ E4} = 204$, $n_{CS\ E4} = 105$. $n_{CSS\ E4} = 231$, $n_{CA\ E5} = 136$, $n_{CS\ E5} = 44$. $n_{CSS\ E5} = 142$. $d_{CS\text{-}CA}$ = effect size for Combat Support-Combat Arms mean difference. $d_{CSS\text{-}CA}$ = effect size for Combat Service Support-Combat Arms mean difference. $d_{CS\text{-}CA}$ = effect size for Combat Support-Combat Service Support mean difference. Effect sizes calculated within pay grade as (mean of 1st MOS type – mean of 2nd MOS type)/Overall *SD*. Overall *SD* = standard deviation calculated across all Soldiers in the given pay grade (regardless of MOS type). Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

At the E5 pay grade, Soldiers in CA MOS viewed themselves as significantly more capable of performing work requiring Leadership Orientation and Stress Tolerance (relative to other types of work) compared to Soldiers in other MOS.

Summary

These results showed that the internal functioning of the predictor measures remained largely the same among the 942 Soldiers participating in the predictor portion of this longitudinal research compared to the 1,889 Soldiers who participated in the concurrent validation effort. That is, subgroup differences on instrument scales, scale reliabilities, and patterns of correlations among the scales within instruments remained remarkably similar. These similarities occurred despite the fact that the mode of administration for the instruments differed across samples. Specifically, in the longitudinal validation sample, all predictor instruments were administered via laptop computer, whereas in the concurrent sample, all instruments were administered via paper-and-pencil.

CHAPTER 4: RESULTS FOR CRITERION DATA COLLECTION INSTRUMENTS

Overview

This chapter documents the results of analyses conducted for responses collected on the NCO Promotion Soldier and supervisor websites. The instruments are discussed one at a time and, given the salience of pay grade differences found in the NCO21 concurrent validation effort (see Knapp, McCloy, & Heffner, 2004), all results are presented by pay grade to the extent possible. For each Soldier instrument, results include:

- Mean score differences across pay grades,
- Internal consistency reliability estimates (where appropriate),
- Correlations among instrument scales, and
- Mean score differences across demographic subgroups (gender, race/ethnicity, and MOS).

For the supervisor instruments, results include:

- Correlations among observed and expected future composite performance rating scales,
- Mean score differences across demographic groups, and
- Means on rating confidence scales.

Soldier Website Data Collection Simulated Promotion Point Worksheet (SimPPW)

The criterion data collection (i.e., Soldier website) version of the PFF21 was scored the same way as the predictor version (see Chapter 3). Scores were generated for (a) Awards, Certificates, and Military Achievements; (b) Military Education; (c) Civilian Education; and (d) Military Training. In addition, a simulated criterion PPW Composite score was calculated for each Soldier by summing the four simulated scores described above. Recall the maximum score that a Soldier could receive on this composite was 500. Note that this maximum score differs from the maximum score on the operational PPW because the simulated PPW does not include Commander's Evaluation points (150) or Promotion Board points (150).

Table 4.1 shows mean Soldier SimPPW scores by pay grade. Like the current validation sample and the predictor data collection values from this research, E5 Soldiers were found to have higher SimPPW scores than E4 Soldiers. However, it is important to note that the Awards, Military Education, and Civilian Education scores were considerably higher than those in the earlier two data collections. This was a reasonable result because this table represents these Soldiers' original pay grade status at the time they began their participation in this research (i.e., during the predictor data collection). In the interim, the Soldiers had substantial opportunity to accumulate additional awards and education. However, the Military Training scores, which are based on physical fitness and weapons tests, were very close to those reported for the earlier two data collections.

Table 4.1. Mean Soldier Website SimPPW Scores by Pay Grade

		E4 Soldiers		E5 Soldiers	
Scale	$d_{E5\text{-}E4}$	*M*	*SD*	*M*	*SD*
Awards	**1.00**	67.32	26.88	94.09	12.89
Military Education	**0.49**	61.08	56.38	88.48	66.77
Military Training	**0.71**	48.65	20.05	62.95	21.54
Civilian Education	**0.58**	25.77	34.79	46.12	42.92
SimPPW Composite	**0.99**	202.83	89.92	291.65	98.08

Note. $n_{E4} = 71$, $n_{E5} = 66$. $d_{E5\text{-}E4}$ = effect size for E5-E4 mean difference. Effect sizes calculated as $(M_{E5} - M_{E4})/SD_{E4}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 4.2 shows correlations among SimPPW scores by pay grade. These results are similar to those in the concurrent validation and predictor sample for this research in terms of which correlations are significant; however, their relative sizes did differ. This finding was likely due to the relatively greater experience of these Soldiers and the possibility that scores on these scales do not change uniformly with tenure. These correlations should be interpreted with some caution given their small sample sizes ($n_{E4} = 71$, $n_{E5} = 66$).

Table 4.2. Soldier Website SimPPW Scale Intercorrelations

Scale	1	2	3	4
E4 Soldiers				
1. Awards				
2. Military Education	.15			
3. Military Training	**.21**	.15		
4. Civilian Education	.08	**.37**	-.02	
5. SimPPW Composite	**.47**	**.85**	**.37**	**.63**
E5 Soldiers				
1. Awards				
2. Military Education	**.24**			
3. Military Training	.02	.16		
4. Civilian Education	-.06	**.31**	.06	
5. SimPPW Composite	**.28**	**.88**	**.36**	**.65**

Note. $n_{E4} = 71$. $n_{E5} = 66$. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

Table 4.3 shows correlations between the predictor and Soldier website versions of the SimPPW scales. Given the relatively small sample sizes for E4 and E5 Soldiers, the correlations for the combined sample were also reported. Correlations between scores on the same scale can be interpreted as the stability of the scale score across time. Correlation in the combined sample showed that Awards and Civilian Education scores were more stable across time than Military Training and SimPPW composite scores, with Military Education being the least stable. Military Education was the least stable across time in all three samples ($r_{E4} = .32$, $r_{E5} = .26$, $r_{Combined} = .33$). A possible explanation is that Military Education may have more to do with MOS membership and unit assignments than the individual's job performance.

Table 4.3. Intercorrelations between Predictor and Soldier Website SimPPW Scales

	Soldier Website Version Scale				
Predictor Version Scale	1	2	3	4	5
E4 Soldiers					
1. Awards	**.57**	.07	.11	-.07	.21
2. Military Education	.02	**.32**	.00	.09	**.23**
3. Military Training	**.24**	**.28**	**.52**	**.23**	**.44**
4. Civilian Education	.04	.04	**.21**	**.51**	.15
5. SimPPW Composite	**.36**	**.28**	.16	**.31**	**.43**
E5 Soldiers					
1. Awards	**.40**	**.30**	-.09	.05	**.26**
2. Military Education	.15	**.26**	.04	.13	**.25**
3. Military Training	.05	-.03	**.43**	-.15	.03
4. Civilian Education	.07	**.25**	.10	**.74**	**.48**
5. SimPPW Composite	**.26**	**.38**	.19	**.50**	**.52**
E4 and E5 Soldiers Combined					
1. Awards	**.64**	**.26**	**.21**	**.15**	**.41**
2. Military Education	**.25**	**.33**	**.14**	**.20**	**.36**
3. Military Training	**.23**	**.16**	**.50**	.08	**.28**
4. Civilian Education	**.17**	**.23**	.09	**.68**	**.44**
5. SimPPW Composite	**.50**	**.39**	**.32**	**.47**	**.59**

Note. $n_{E4} = 71$. $n_{E5} = 66$. $n_{Combined} = 137$. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

Table 4.4 shows mean SimPPW scores by gender. Female-male effect sizes were not calculated because neither the E4 nor E5 samples included a minimum sample size of 20. Although the relative sizes of the female means showed some variation from the predictor data collection (especially for E5s), that is likely due to small sample sizes. Otherwise, these means were similar to the predictor data collection values.

Table 4.4. Mean Soldier Website SimPPW Scores by Gender

	Male		Female	
Scale	*M*	*SD*	*M*	*SD*
E4 Soldiers				
Awards	68.55	27.19	63.13	26.20
Military Education	59.75	54.75	65.69	63.33
Military Training	50.80	20.45	41.25	17.17
Civilian Education	24.44	36.23	30.38	29.90
SimPPW Composite	203.53	93.24	200.44	80.15
E5 Soldiers				
Awards	93.33	13.67	98.89	3.33
Military Education	88.14	67.87	90.67	63.03
Military Training	67.25	19.48	35.78	12.23
Civilian Education	43.84	42.19	60.56	47.29
SimPPW Composite	292.56	99.72	285.89	92.18

Note. $n_{Male\ E4} = 55$, $n_{Female\ E4} = 16$. $n_{Male\ E5} = 57$, $n_{Female\ E5} = 9$.

Table 4.5 shows mean SimPPW scores by race/ethnicity. Hispanic-White effect sizes were not calculated because neither the E4 nor E5 samples included a minimum sample size of 20. The relative sizes in means across groups are not particularly consistent with either of the last two data collections. This result is likely due to the relatively small number of Soldiers in all three subgroups of interest.

Table 4.5. Mean Website SimPPW Scores by Race/Ethnic Group

		White		Black		Hispanic	
Scale	$d_{B\text{-}W}$	*M*	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers							
Awards	0.71	61.73	26.38	80.59	20.14	55.80	33.64
Military Education	-0.35	69.49	57.35	49.36	56.57	48.40	54.01
Military Training	0.24	44.59	19.64	49.32	20.50	61.70	18.12
Civilian Education	-0.03	27.65	34.37	26.50	37.37	21.40	35.95
SimPPW Composite	0.03	203.46	91.77	205.77	84.09	187.30	109.10
E5 Soldiers							
Awards		94.88	12.33	90.77	16.18	92.78	13.25
Military Education		87.63	66.16	76.69	63.48	109.56	73.98
Military Training		64.28	21.84	57.92	15.89	71.67	27.15
Civilian Education		45.95	42.71	51.00	45.64	49.22	41.75
SimPPW Composite		292.73	94.01	276.38	98.94	323.22	113.82

Note. $n_{\text{White E4}} = 37$, $n_{\text{Black E4}} = 22$. $n_{\text{Hispanic E4}} = 10$, $n_{\text{White E5}} = 40$, $n_{\text{Black E5}} = 13$. $n_{\text{Hispanic E5}} = 9$. $d_{B\text{-}W}$ = effect size for Black-White mean difference. Effect sizes calculated within pay grade as $(M_{\text{Black}} - M_{\text{White}})/SD_{\text{White}}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 4.6 shows mean SimPPW scores by MOS type. CSS-CA effect sizes for E4 Soldiers were the only ones calculated because samples for subgroups in the other comparisons did not included a minimum sample size of 20. The relative sizes in means across groups are not

Table 4.6. Mean Soldier Website SimPPW Scores by MOS Type

		Combat Arms		Combat Support		Combat Service Support	
Scale	$d_{CSS\text{-}CA}$	*M*	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers							
PPW Awards	-0.11	69.30	24.42	65.80	28.75	66.73	28.16
PPW Military Education	-0.08	68.00	59.51	38.60	37.22	63.20	58.42
PPW Military Training	0.18	46.25	22.40	46.90	15.67	50.24	20.09
PPW Civilian Education	**0.74**	13.70	26.87	17.80	30.02	33.61	37.69
Simulated PPW Composite	0.19	197.25	88.57	169.10	59.17	213.78	95.98
E5 Soldiers							
PPW Awards		95.53	9.99	94.33	12.08	93.13	14.91
PPW Military Education		87.05	64.99	84.27	71.73	91.31	67.48
PPW Military Training		59.58	21.62	66.40	25.35	63.34	19.95
PPW Civilian Education		48.32	44.78	44.73	42.92	45.47	43.16
Simulated PPW Composite		290.47	100.69	289.73	107.95	293.25	94.95

Note. $n_{\text{CA E4}} = 20$, $n_{\text{CS E4}} = 10$. $n_{\text{CSS E4}} = 41$, $n_{\text{CA E5}} = 19$, $n_{\text{CS E5}} = 15$. $n_{\text{CSS E5}} = 32$. $d_{CSS\text{-}CA}$ = effect size for Combat Service Support-Combat Arms mean difference. Effect sizes calculated within pay grade as (mean of 1st MOS type – mean of 2nd MOS type)/Overall *SD*. Overall *SD* = standard deviation calculated across all Soldiers in the given pay grade (regardless of MOS type). Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

particularly consistent with either of the last two data collections. This result is likely due to the relatively small number of Soldiers in all three subgroups of interest. However, it is interesting to note that for these Soldier website scores, CSS Soldiers had a mean Civilian Education score almost ¾ of an *SD* greater and CA Soldiers. This comparison showed the largest effect size in the full predictor sample for E5 Soldiers in this research ($d_{CSS\text{-}CA} = 0.82$). However, for E4 Soldiers in the predictor sample, the effect size was smaller ($d_{CSS\text{-}CA} = 0.12$). This pattern of findings is consistent with the possibility that CA Soldiers had less opportunity to pursue civilian education than CSS Soldiers. As the E4 Solider predictor sample effect size shows, the effect of this possible differential opportunity may expand over time.

Soldier Website Data Collection Experience and Activities Record (ExAct)

The criterion data collection (i.e., Soldier website) version of the ExAct was scored exactly the same way as the predictor version (see Chapter 3). Scores were generated for (a) Computer Experience, (b) Supervisory Experience, and (c) General Experience.

Table 4.7 shows mean ExAct scores by pay grade. In this sample the effect sizes showed the same relative size across scales as they did in the predictor sample favoring E5 Soldiers (i.e., General Experience being the largest and Computer Experience being the smallest). It is possible that differences in computer experience become smaller as all Soldiers got more experience. The manner in which the scores were standardized could mask this effect (see Chapter 3 for discussion of standardization).

Table 4.7. Mean Soldier Website ExAct Scores by Pay Grade

		E4 Soldiers		E5 Soldiers	
Scale	$d_{E5\text{-}E4}$	*M*	*SD*	*M*	*SD*
Computer Experience	0.00	0.00	0.56	0.00	0.62
Supervisory Experience	**0.79**	-0.29	0.76	0.31	0.41
General Experience	**1.23**	-0.26	0.44	0.28	0.32

Note. $n_{E4} = 70$, $n_{E5} = 65$. $d_{E5\text{-}E4}$ = effect size for E5-E4 mean difference. Effect sizes calculated as $(M_{E5} - M_{E4})/SD_{E4}$. Statistically significant effect sizes are bolded, $p < .05$ (two-tailed).

Table 4.8 shows correlations among ExAct scores by pay grade, as well as internal consistency reliability estimates for each scale. Generally, these results were similar to those in the predictor data collection sample for this project. The primary difference is that the correlation between Computer and General Experience among E5 Soldiers is far higher in this sample (.55) than in the predictor sample (.24).

Table 4.9 shows correlations between the predictor and Soldier website versions of the ExAct scales. Again, given the relatively small sample sizes for E4 and E5 Soldiers, the correlations for the combined sample also were reported. Correlations in the combined sample showed that General Experience scores were the most stable across time, whereas Computer Experience scores were the least stable.

Table 4.8. Soldier Website ExAct Scale Intercorrelations and Reliability Estimates

Scale	1	2	3
E4 Soldiers			
1. Computer Experience	(0.73)		
2. Supervisory Experience	**0.31**	(0.90)	
3. General Experience	**0.45**	**0.66**	(0.85)
E5 Soldiers			
1. Computer Experience	(0.74)		
2. Supervisory Experience	0.18	(0.79)	
3. General Experience	**0.55**	**0.44**	(0.71)

Note. $n_{E4} = 70$. $n_{E5} = 65$. Internal consistency reliability estimates (alpha) are shown in parentheses on the diagonal. Bolded correlations are statistically significant, $p < .05$ (one-tailed).

Table 4.9. Intercorrelations between Predictor and Soldier Website ExAct Scales

	Soldier Website Version Scale		
Predictor Version Scale	1	2	3
E4 Soldiers			
1. Computer Experience	**0.44**	-0.01	0.07
2. Supervisory Experience	**0.24**	**0.53**	**0.53**
3. General Experience	**0.29**	**0.43**	**0.73**
E5 Soldiers			
1. Computer Experience	**0.42**	-0.10	0.17
2. Supervisory Experience	0.09	**0.50**	**0.34**
3. General Experience	**0.28**	**0.32**	**0.52**
E4 and E5 Soldiers Combined			
1. Computer Experience	**0.42**	0.05	**0.20**
2. Supervisory Experience	0.13	**0.65**	**0.67**
3. General Experience	**0.22**	**0.56**	**0.78**

Note. $n_{E4} = 70$. $n_{E5} = 65$. $n_{E5} = 135$. Bolded correlations are statistically significant, $p < .05$ (one-tailed).

Table 4.10 shows mean ExAct score by gender for each pay grade. Female-male effect sizes were not calculated because neither the E4 nor E5 samples included a minimum sample size of 20. However, the relative sizes in mean values comparing gender are similar across ExAct scales compared to the predictor data collection values.

Table 4.10. Mean Soldier Website ExAct Scores by Gender

	Male		Female	
Scale	*M*	*SD*	*M*	*SD*
E4 Soldiers				
Computer Experience	-0.03	0.58	0.10	0.48
Supervisory Experience	-0.28	0.78	-0.32	0.71
General Experience	-0.22	0.48	-0.40	0.29
E5 Soldiers				
Computer Experience	-0.01	0.65	0.05	0.49
Supervisory Experience	0.34	0.39	0.15	0.51
General Experience	0.31	0.33	0.08	0.21

Note. $n_{Male\ E4} = 54$, $n_{Female\ E4} = 16$. $n_{Male\ E5} = 56$, $n_{Female\ E5} = 9$.

Table 4.11 shows mean Soldier website scores by race/ethnicity. Hispanic-White effect sizes were not calculated because neither the E4 nor E5 samples included a minimum sample size of 20. Note that none of the calculated effect sizes were significant. Similar to results for the Soldier website ExAct scores, the relative sizes in means across groups are not particularly consistent with either of the last two data collections. This result is likely due to the relatively small number of Soldiers in all three subgroups of interest.

Table 4.11. Mean Website ExAct Scores by Race/Ethnic Group

		White		Black		Hispanic	
Scale	$d_{B\text{-}W}$	*M*	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers							
Computer Experience	-0.70	0.00	0.56	-0.39	0.77	-0.37	0.41
Supervisory Experience	-0.57	0.09	0.53	-0.22	0.80	-0.11	0.46
General Experience	0.20	-0.17	0.70	-0.03	0.68	-0.22	0.50
E5 Soldiers							
Computer Experience		0.05	0.59	0.29	0.45	0.28	0.33
Supervisory Experience		0.19	0.32	0.42	0.33	0.36	0.27
General Experience		-0.02	0.56	0.31	0.39	0.30	0.31

Note. $n_{White\ E4} = 37$, $n_{Black\ E4} = 22$. $n_{Hispanic\ E4} = 9$, $n_{White\ E5} = 39$, $n_{Black\ E5} = 13$. $n_{Hispanic\ E5} = 9$. $d_{B\text{-}W}$ = effect size for Black-White mean difference. Effect sizes calculated within pay grade as $(M_{Black} - M_{White})/SD_{White}$. No statistically significant effect sizes were found.

Table 4.12 shows mean ExAct scores by MOS type. No effect sizes were calculated because neither subgroup included a minimum sample size of 20 for any comparison. The relative sizes in means across groups are not particularly consistent with either of the last two data collections. This result is likely due to the relatively small number of Soldiers in all three subgroups of interest.

Table 4.12. Mean Soldier Website ExAct Scores by MOS Type

	Combat Arms		Combat Support		Combat Service Support	
Scale	*M*	*SD*	*M*	*SD*	*M*	*SD*
E4 Soldiers						
ExAct Computer Experience	0.04	0.64	-0.19	0.76	-0.21	0.42
ExAct Supervisory Experience	0.01	0.39	-0.36	0.71	-0.32	0.38
ExAct General Experience	-0.02	0.56	-0.32	0.79	-0.27	0.48
E5 Soldiers						
ExAct Computer Experience	0.03	0.46	0.22	0.51	0.28	0.32
ExAct Supervisory Experience	0.16	0.37	0.39	0.38	0.32	0.30
ExAct General Experience	-0.10	0.79	0.34	0.35	0.27	0.34

Note. $n_{CA\ E4} = 19$, $n_{CS\ E4} = 10$. $n_{CSS\ E4} = 41$, $n_{CA\ E5} = 19$, $n_{CS\ E5} = 15$. $n_{CSS\ E5} = 31$.

Criterion Data Collection Supervisor Ratings

As described in Chapter 2, supervisors rated Soldiers on 21 observed performance scales (i.e., 19 dimensions of observed performance, an overall effectiveness scale, and a senior NCO potential scale) and 6 expected future performance scales describing conditions NCOs are likely to face in the future Army. Additionally, raters evaluated their overall confidence regarding their observed performance ratings and their confidence relative to each of their six expected future performance ratings. Here, we present descriptive statistics for the (a) observed performance composite (without Scale 17 [Coordinating Multiple Units and Battlefield Functions] because too few Supervisors made ratings on this scale), (b) overall effectiveness, (c) senior NCO potential, (d) expected future performance composite, and (e) confidence ratings. The results for performance ratings are not reported by pay grade for correlations and most of the subgroup analyses because of the small number of Soldiers with usable ratings (total $n = 53$ for observed ratings; total $n = 56$ for expected future performance ratings).

The mean confidence rating for the observed performance ratings was 5.94 on a 7-point scale ($SD = 0.89$). A composite confidence rating was calculated for each rater consisting of the mean across the six 7-point-scale confidence ratings for each expected future performance scale. The mean of this composite was 5.68 ($SD = 1.38$).

Table 4.13 shows the mean supervisor rating scores by pay grade. These effect sizes cannot easily be compared to those in the concurrent validation sample because we collected performance ratings only for current E5 and E6 Soldiers. However, the E5 Soldier means from the earlier concurrent validity effort are close to those shown in Table 4.13 (e.g., $M = 5.03$ and $M = 4.86$ for the Observed and Expected Future Performance Composites, respectively).

Table 4.13. Mean Supervisor Performance Rating Scores by Gender

		E4 Soldiers		E5 Soldiers	
Scale	$d_{E5\text{-}E4}$	*M*	*SD*	*M*	*SD*
Observed Performance Composite	-0.17	5.09	0.66	4.98	1.04
Overall Effectiveness Rating Scale	-0.17	5.36	0.94	5.20	1.34
Senior NCO Potential Rating Scale	-0.02	4.91	1.18	4.88	1.79
Expected Future Performance Composite	-0.30	5.11	0.99	4.82	1.44

Note. $n_{E4} = 29\text{-}32$, $n_{E5} = 24\text{-}25$. $d_{E5\text{-}E4}$ = effect size for E5-E4 mean difference. Effect sizes calculated as $(M_{E5} - M_{E4})/SD_{E4}$. None of the effect sizes were found to be significant.

Table 4.14 shows correlations among relevant ratings scores. The correlations among observed scores in this project were very similar to those in the concurrent validation effort. For example, the correlation between the Observed Performance Composite and Overall Effectiveness rating scale in the concurrent validation was .84 for E5 Soldiers. The correlations between the Observed and Expected Future Performance Composites were also very similar for both efforts (i.e., E5 Soldiers in concurrent validation – $r_{\text{Observed, Expected Future}} = .81$).

Table 4.14. Intercorrelations Among Supervisor Performance Rating Scores

Scale	1	2	3
1. Observed Performance Composite			
2. Overall Effectiveness Rating Scale	**.87**		
3. Senior NCO Potential Rating Scale	**.80**	**.87**	
4. Expected Future Performance Composite	**.75**	**.81**	**.81**

Note. n = 53-56. Bolded correlations are statistically significant, $p < .05$ (one-tailed).

The concurrent validation effort had a sufficient number of E5 and E6 Soldiers who each had two supervisor ratings to support the calculation of interrater reliability estimates (Sager, Putka, & McCloy, 2004). The single-rater interrater reliability estimates for E5 Soldiers for the Observed and Expected Future Performance were .45 and .31, respectively. These values were used in this research to calculate weighted interrater reliability estimates that were used to correct criterion-related validity estimates for criterion unreliability.[17]

Table 4.15 shows mean ratings scores by gender. Effect sizes were not calculated because neither subgroup included a minimum sample size of 20. However, the relative sizes of the means are similar to those in the concurrent validation sample. The male and female concurrent validation means for the Observed Composite were 5.05 and 4.89, respectively; for the Expected Future Performance Composite they were 4.91 and 4.53, respectively.

Table 4.15. Mean Supervisor Performance Rating Scores by Gender

	Male		Female	
Scale	*M*	*SD*	*M*	*SD*
1. Observed Performance Composite	5.10	0.91	4.80	0.35
2. Overall Effectiveness Rating	5.42	1.16	4.79	0.78
3. Senior NCO Potential Rating	5.08	0.15	4.21	1.23
4. Expected Future Performance Composite	5.08	1.24	4.61	1.08

Note. n_{Male} = 42-46, n_{Female} = 11-12.

Table 4.16 shows mean rating scores by race/ethnic group. Effect sizes were not calculated because neither subgroup included a minimum sample size of 20. Similar to the concurrent validation sample, the White and Black mean ratings showed no or very small difference. The concurrent validation sample did not have a sufficient number of Hispanics to report their means, and a sample size of n = 7 Hispanics in this research is too small to interpret their means.

[17] The interrater reliability estimates used to correct the validity estimates were weighted based on the number of Soldiers whose composite scores were based on ratings provided by one or two supervisors. For the Observed Performance Composite, 45 Soldiers were rate by one supervisor and 11 were rated by two supervisors. For the Expected Future Composite, 42 Soldiers were rated by one supervisor and 11 were rated by two supervisors.

Table 4.16. Mean Supervisor Performance Rating Scores by Race/Ethnic Group

	White		Black		Hispanic	
Scale	*M*	*SD*	*M*	*SD*	*M*	*SD*
1. Observed Performance Composite	5.06	0.85	5.06	0.72	5.02	1.13
2. Overall Effectiveness Rating	5.33	1.16	5.21	0.99	5.29	1.38
3. Senior NCO Potential Rating	4.85	1.46	4.79	1.61	5.36	1.38
4. Expected Future Performance Composite	4.91	1.33	5.03	1.11	5.26	0.98

Note. n_{White} = 33-36. n_{Black} = 12-14. $n_{Hispanic}$ = 7.

Summary

Predictor and Soldier website versions of the PFF21 and ExAct did perform similarly. These instruments were completed by Soldiers between 14 and 19 months apart. The contents of the instruments were identical. However, the predictor versions were administered to groups of Soldiers via laptop computers and supervised by test administrators, whereas for the next version Soldiers logged on to the NCO Promotion Soldier website and completed the instruments on their own. Comparisons of subgroup differences and correlations among scales, within and across versions, suggest that instruments functioned similarly across time and modes of administration. Correlations between the same scales across occasions did not show the level of stability that would be expected for test-retest reliabilities of trait measures (e.g., general cognitive ability). Given that these are measures of experience, however, we judge them to be reasonable.

The sample size of Soldiers with job performance ratings was not sufficient to perform all of the planned analyses. However, to the extent that comparisons were possible, subgroup differences and relations among rating scales and composites for this project's sample were remarkably similar to results in the concurrent validation sample. This finding offers some evidence to support the construct validity of the observed and expected future performance scales.

CHAPTER 5: CROSS-INSTRUMENT ANALYSES

Overview

The two previous chapters focused on providing results regarding each instrument individually. In this chapter, we provide results regarding interrelations among instruments. First, we present correlations among scales for only those instruments that were administered during the predictor data collection. Next, we present correlations between scales for those original LAT instruments and scales from the WSI. This analysis is followed by describing the relations among the instruments administered to Soldiers during the criterion data collection (i.e., Soldier website versions of the PFF21 and ExAct). These correlations among predictor instruments are followed by validity results comparing the predictors to the job performance ratings criteria and the promotion criterion.

Relations Among Predictors

To facilitate comparisons with cross-instrument tables presented in the concurrent validation report (Knapp, et al., 2004), we have presented the scales in Tables 5.1 and 5.2 in the same order they appear in the concurrent validation report. Scores on instruments designed to assess cognitive aptitude and skills related to judgment are shown first (i.e., ASVAB and LeadEx), followed by instruments emphasizing experience (i.e., SimPPW and ExAct) and, lastly, instruments designed to assess temperament constructs (i.e., SDI and IQ-II).

Cognitive Aptitude and Judgment

ASVAB

The ASVAB GT composite score is currently used for various post-enlistment decisions (e.g., eligibility for reenlistment) and can be considered a good measure of general cognitive aptitude. As in the concurrent validation sample, ASVAB GT was most related to the LeadEx composite and IQ-II Tolerance for Ambiguity scale.

LeadEx

Among E4 Soldiers, the strongest correlate of the LeadEx was ASVAB GT, followed by several temperament variables (e.g., SDI Work Orientation, Dependability, and Agreeableness). Generally, correlations between temperament variables and LeadEx scores were weaker here than they were in the concurrent validation sample for E4 Soldiers. Indeed, in the concurrent validation sample, temperament variables were the strongest correlates of the LeadEx among E4 Soldiers, with many correlations in the mid .20s to mid .30s. In this research they are in the .10s to mid .20s. Similar to the concurrent validation results, among E5 Soldiers, the strongest predictors of the LeadEx composites were temperament variables followed by the ASVAB GT. In general the patterns of correlations in both pay grades were very similar to those found in the concurrent validation effort. Specifically, LeadEx scores were significantly related to almost all of the SDI and IQ-II scales. Such findings suggest that personality influences Soldiers' evaluations of the best and worst ways to behave in different situations.

Table 5.1. Intercorrelations among Original Leadership Assessment Tool Scales for E4 Soldiers

Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. EMF: ASVAB GT Score																						
2. LeadEx: 24-Item Composite	**.30**																					
3. LeadEx: 40-Item Composite	**.29**	**.94**																				
4. Simulated PPW Composite	-.01	.03	.04																			
5. PPW Awards	-.04	.03	.03	**.64**																		
6. PPW Military Education	.02	.04	.02	**.56**	**.16**																	
7. PPW Civilian Education	.05	.02	.06	**.52**	.04	**.08**																
8. PPW Military Training	-.05	-.02	-.04	**.55**	**.13**	**.15**	.05															
9. ExAct: Computer Experience	**.12**	**.11**	**.13**	**.13**	.05	**.07**	**.10**	.07														
10. ExAct: Supervisory Experience	**-.09**	**.11**	**.11**	**.30**	**.22**	**.22**	-.03	**.29**	**.30**													
11. ExAct: General Experience	.00	**.13**	**.13**	**.42**	**.33**	**.21**	**.09**	**.32**	**.31**	**.70**												
12. SDI: Dependability	-.03	**.23**	**.26**	**.07**	**.08**	.03	.00	.05	**.15**	**.14**	**.10**											
13. SDI: Adjustment	**.07**	**.10**	**.10**	**.10**	.01	.04	.01	**.19**	**.10**	**.16**	**.14**	**.32**										
14. SDI: Work Orientation	-.04	**.25**	**.26**	**.24**	**.13**	**.13**	**.08**	**.22**	**.19**	**.35**	**.28**	**.45**	**.36**									
15. SDI: Agreeableness	.03	**.22**	**.23**	.06	.03	.00	.01	**.10**	**.10**	**.09**	**.08**	**.55**	**.51**	**.43**								
16. SDI: Physical Conditioning	-.04	**.09**	**.08**	**.15**	.04	.05	.03	**.25**	.04	.05	-.01	**.29**	**.31**	**.46**	**.35**							
17. SDI: Leadership	.04	**.20**	**.20**	**.22**	**.12**	**.09**	**.08**	**.21**	**.18**	**.37**	**.32**	**.28**	**.43**	**.63**	**.21**	**.12**						
18. IQ-II: Hostility to Authority	-.06	**-.17**	**-.23**	-.02	.00	-.03	-.04	.02	-.06	-.02	.03	**-.44**	**-.23**	**-.28**	**-.36**	**-.18**	**-.18**					
19. IQ-II: Manipulativeness	**-.09**	**-.23**	**-.25**	**-.08**	-.04	-.04	-.02	**-.09**	-.04	**-.13**	**-.08**	**-.45**	**-.34**	**-.45**	**-.42**	**-.22**	**-.37**	**.51**				
20. IQ-II: Social Perceptiveness	**.12**	**.12**	**.11**	**.13**	.05	.05	.03	**.16**	**.21**	**.25**	**.32**	**.08**	**.27**	**.29**	**.10**	.03	**.45**	**.15**	**-.15**			
21. IQ-II: Tolerance for Ambiguity	**.19**	**.20**	**.21**	**.12**	**.09**	.03	**.08**	**.07**	**.20**	**.20**	**.25**	**.17**	**.38**	**.33**	**.17**	.04	**.46**	**-.20**	**-.34**	**.37**		
22. IQ-II: Emergent Leadership	**.07**	**.15**	**.14**	**.27**	**.17**	**.14**	.06	**.26**	**.31**	**.47**	**.48**	**.22**	**.35**	**.48**	**.15**	**.14**	**.69**	.01	**-.28**	**.66**	**.46**	
23. IQ-II: Interpersonal Skills	**.09**	**.21**	**.23**	.06	.03	-.01	-.02	**.13**	**.18**	**.22**	**.23**	**.45**	**.39**	**.36**	**.53**	**.18**	**.33**	**-.40**	**-.44**	**.24**	**.33**	**.35**

Note. n = 540-591. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

Table 5.2. Intercorrelations among Original Leadership Assessment Tool Scales for E5 Soldiers

Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. EMF: ASVAB GT Score																						
2. LeadEx: 24-Item Composite	**.19**																					
3. LeadEx: 40-Item Composite	**.19**	**.93**																				
4. Simulated PPW Composite	-.08	.06	.08																			
5. PPW Awards	-.05	-.03	-.03	**.46**																		
6. PPW Military Education	**-.17**	.03	.04	**.67**	**.14**																	
7. PPW Civilian Education	.01	.05	.08	**.64**	.08	**.11**																
8. PPW Military Training	.08	.07	.06	**.30**	.02	-.03	-.06															
9. ExAct: Computer Experience	.07	**.11**	**.12**	**.25**	**.14**	.08	**.27**	.00														
10. ExAct: Supervisory Experience	.00	.05	.04	**.17**	**.21**	.07	.00	**.19**	**.13**													
11. ExAct: General Experience	.03	**.12**	**.11**	**.27**	**.35**	.05	.06	**.27**	**.24**	**.56**												
12. SDI: Dependability	.01	**.29**	**.31**	**.11**	.04	.01	**.14**	.03	.07	.07	.08											
13. SDI: Adjustment	**.19**	**.22**	**.21**	.09	.01	-.02	.05	**.20**	.07	**.10**	**.20**	**.36**										
14. SDI: Work Orientation	**.12**	**.24**	**.24**	**.10**	-.01	.01	.00	**.29**	**.11**	**.38**	**.31**	**.29**	**.28**									
15. SDI: Agreeableness	.02	**.24**	**.21**	.02	.00	-.01	.05	.01	.05	**.14**	**.09**	**.45**	**.47**	**.35**								
16. SDI: Physical Conditioning	**.14**	.01	.01	.06	-.07	.03	-.01	**.21**	.02	.08	.09	**.20**	**.20**	**.35**	**.27**							
17. SDI: Leadership	**.19**	**.31**	**.30**	.05	-.05	-.06	-.02	**.34**	**.09**	**.33**	**.33**	**.19**	**.35**	**.63**	**.17**	.05						
18. IQ-II: Hostility to Authority	-.08	**-.22**	**-.22**	-.01	.06	-.04	-.02	.00	-.06	-.03	-.04	**-.35**	**-.37**	**-.31**	**-.42**	**-.10**	**-.22**					
19. IQ-II: Manipulativeness	-.05	**-.25**	**-.25**	-.08	.04	-.04	-.05	**-.12**	-.01	**-.19**	**-.21**	**-.26**	**-.34**	**-.42**	**-.40**	**-.15**	**-.38**	**.54**				
20. IQ-II: Social Perceptiveness	**.10**	**.16**	**.16**	**.13**	.09	-.01	.03	**.26**	**.15**	**.27**	**.32**	**.09**	**.17**	**.38**	**.17**	**.19**	**.38**	.02	**-.22**			
21. IQ-II: Tolerance for Ambiguity	**.24**	**.20**	**.22**	-.01	-.04	-.04	-.04	**.16**	**.14**	**.25**	**.27**	.03	**.30**	**.36**	.06	.05	**.44**	**-.21**	**-.32**	**.32**		
22. IQ-II: Emergent Leadership	**.15**	**.25**	**.21**	**.10**	.03	-.06	.00	**.34**	**.23**	**.43**	**.45**	**.13**	**.23**	**.52**	**.13**	**.12**	**.64**	-.07	**-.29**	**.60**	**.44**	
23. IQ-II: Interpersonal Skills	**.17**	**.29**	**.29**	.00	-.07	-.02	-.01	**.14**	**.10**	.08	**.12**	**.22**	**.39**	**.33**	**.50**	**.13**	**.31**	**-.46**	**-.47**	**.32**	**.32**	**.33**

Note. $n = 332\text{-}351$. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

Experience-Oriented Measures

SimPPW

Aside from correlations with the overall SimPPW Composite, the SimPPW Awards and Military Education scores correlated most highly with the ExAct General Experience and Supervisory Experience scales for both E4 and E5 Soldiers. As in the concurrent validation sample, correlations between the SimPPW Civilian Education scores and other non-PPW scores were generally small. The only exception was a correlation of .27 between SimPPW Civilian Education and Computer Experience for E5 Soldiers.

Similar to the concurrent validation sample, several variables were significantly correlated with SimPPW Military Training. Recall that SimPPW Military Training reflects Soldiers' scores on the APFT and a weapons qualification test. Among the variables most related to Soldiers' performance on these tests were IQ-II Emergent Leadership, SDI Leadership, SDI Physical Conditioning, SDI Work Orientation, and the ExAct's Supervisory and General Experience scores.

ExAct

Among Soldiers at both pay grades, the strongest correlates of the ExAct experience scores (particularly Supervisory and General Experience) were the SDI and IQ-II Leadership scores, SDI work orientation, IQ-II Social Perceptiveness, and SimPPW Awards. These findings are similar to results from the concurrent validation effort.

Temperament Measures

SDI and IQ-II

The above sections have addressed relations between the temperament measures and other instruments. In this section, we focus on interrelations between the SDI and IQ-II. Examining correlations between SDI and IQ-II scales reveals evidence for both discriminant and convergent validity of the measures. For example, evidence for the discriminant validity of the measures is apparent in the generally low to moderate correlations among scales from the two instruments. Specifically, none of the SDI scales is so highly correlated with IQ-II scales that it would suggest the two measures are redundant, or that they are failing to offer different perspectives on the temperament of individual Soldiers. Conversely, evidence for the convergent validity is apparent for many of these instruments' scales. The highest correlations between SDI and IQ-II scales were found among those scales that are most conceptually related. For example, the SDI Leadership and IQ-II Emergent Leadership scales were correlated .69 for E4 Soldiers and .64 for E5 Soldiers. The strongest correlate of IQ-II Interpersonal Skills was SDI Agreeableness. Overall, these findings are similar to results from the concurrent validation effort.

WSI

Given the WSI was not administered as part of the concurrent validation effort, we chose to address its relation to other instruments separately. Tables 5.3 and 5.4 show correlations between the WSI scores and original LAT instrument scales for E4 and E5 Soldiers, respectively.

Given the ipsative nature of the WSI scores, caution should be taken in interpreting these correlations. Specifically, positive correlations between WSI scores and other variables indicate that Soldiers who have high standing on a given variable tended to view themselves as more capable of performing a given type of work (linked to a particular trait), relative to other types of work. Conversely, negative correlations between WSI scores and other variables indicate that Soldiers who have high standing on a given variable tended to view themselves as less capable of performing a given type of work (again, linked to a particular trait), relative to other types of work.

Given this context, one can begin to meaningfully interpret the WSI correlations with the other scales. For example, Soldiers who scored high on ExAct Supervisory Experience and SDI Work Orientation tended to view themselves as more capable of performing work requiring Achievement/Effort compared to other types of work.

Table 5.3. Intercorrelations between WSI and Original Leadership Assessment Scales for E4 Soldiers

Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
EMF: ASVAB GT Score	-.05	-.02	-.07	**-.19**	**-.12**	**-.11**	**-.08**	**.23**	.04	**.16**	-.02	.01	.00	-.01	**.16**	.04
LeadEx: 24-Item Composite	.01	.00	.01	**-.08**	**-.09**	-.02	**-.10**	.01	**.09**	**.15**	.04	-.04	-.03	.00	-.01	.03
LeadEx: 40-Item Composite	-.01	.02	.03	-.04	-.07	-.03	**-.11**	.02	**.08**	**.12**	.03	**-.07**	-.04	.02	-.01	.05
Simulated PPW Composite	**.10**	-.02	.04	**-.13**	-.07	**.12**	**.08**	-.02	-.06	.02	**.09**	-.03	.02	-.03	-.02	**-.08**
PPW Awards	.02	-.01	**.09**	**-.15**	**-.09**	**.14**	-.03	.02	-.06	.00	**.09**	.05	.05	-.05	.02	**-.08**
PPW Military Education	**.09**	.06	-.01	-.05	.00	.04	.00	-.05	.02	.06	-.02	-.04	-.03	.00	**-.08**	.02
PPW Civilian Education	**.07**	.01	.04	.02	.00	.03	.03	.03	-.02	-.02	-.01	**-.09**	-.04	.03	**-.08**	-.02
PPW Military Training	.06	**-.09**	-.07	**-.09**	-.05	.05	**.20**	**-.07**	-.06	.02	**.13**	-.01	.06	-.05	.07	**-.08**
ExAct: Computer Experience	.02	-.01	.04	-.05	**-.10**	-.07	.00	.01	.01	-.02	.06	-.05	.02	-.02	**.11**	.04
ExAct: Supervisory Experience	**.16**	.01	**.07**	**-.08**	**-.09**	.06	.06	**-.08**	.03	-.05	**.22**	-.03	.00	**-.18**	.01	**-.10**
ExAct: General Experience	.07	.01	.05	**-.13**	**-.10**	.01	-.01	.00	.07	.03	**.17**	.00	.02	**-.20**	**.11**	-.07
SDI: Dependability	**.13**	.03	**.14**	.06	.05	.00	-.07	-.05	-.04	**-.08**	.05	**-.09**	.01	-.01	**-.16**	.00
SDI: Adjustment	.02	.05	-.05	-.06	-.04	**-.09**	.02	**-.08**	-.05	**.09**	**.11**	-.03	.07	-.04	.05	.06
SDI: Work Orientation	**.23**	-.05	**.13**	**-.13**	**-.13**	.03	**.08**	**-.12**	.01	-.07	**.21**	-.03	.01	-.06	-.02	-.06
SDI: Agreeableness	**.08**	.04	**.10**	.06	.06	-.06	-.06	**-.09**	-.01	-.07	-.03	-.07	**.10**	.00	-.07	.04
SDI: Physical Conditioning	**.11**	-.01	**.11**	.00	.00	.01	**.18**	**-.11**	**-.07**	**-.07**	-.01	-.04	.03	.04	**-.12**	-.03
SDI: Leadership	**.12**	**-.10**	-.04	**-.17**	**-.10**	.01	.03	**-.11**	**.08**	.06	**.39**	**-.10**	-.01	-.01	.05	**-.09**
IQ-II: Hostility to Authority	-.05	.01	**-.07**	.00	-.04	-.05	.03	.05	-.05	.04	.03	**.12**	-.01	.00	.06	-.07
IQ-II: Manipulativeness	**-.15**	.05	**-.08**	.06	.06	-.05	.05	**.08**	.02	.00	**-.09**	.07	.00	.01	.01	-.05
IQ-II: Social Perceptiveness	-.01	.00	**-.08**	**-.08**	**-.12**	**-.10**	.03	-.03	-.07	**.19**	**.22**	-.04	.03	**-.11**	**.15**	.02
IQ-II: Tolerance for Ambiguity	**.07**	.01	**-.12**	**-.19**	**-.14**	**-.08**	-.05	.05	**.09**	**.17**	**.17**	-.01	-.04	**-.10**	**.16**	.01
IQ-II: Emergent Leadership	.07	-.02	-.07	**-.14**	**-.15**	-.05	.04	-.02	.01	**.11**	**.35**	-.06	.01	**-.09**	**.11**	**-.08**
IQ-II: Interpersonal Skills	**.08**	.06	-.03	.00	.04	**-.09**	.00	**-.08**	.06	.02	.05	**-.10**	.02	-.01	.01	-.01

Note. n = 504-540. Statistically significant correlations are bolded, $p < .05$ (one-tailed). Each column corresponds to a different WSI scale: 1 = Achievement/Effort; 2 = Adaptability/Flexibility; 3 = Attention to Detail; 4 = Concern for Others; 5 = Cooperation; 6 = Dependability; 7 = Energy; 8 = Independence; 9 = Initiative; 10 = Innovation; 11 = Leadership Orientation; 12 = Persistence; 13 = Self-Control; 14 = Social Orientation; 15 = Stress Tolerance; 16 = Cultural Tolerance.

Soldiers who scored high on ASVAB GT, ExAct General Experience, IQ-II Tolerance for Ambiguity, and SDI and IQ-II Leadership scales tended to view themselves as less capable of performing work requiring Concern for Others compared to other types of work. A similar (yet weaker) pattern of findings was found with regard to work requiring Cooperation. These findings suggest that Soldiers, who may be good at the "initiating structure" aspects of leadership, view themselves as less capable of performing the "consideration" aspects of leadership.

Soldiers who scored high on SDI Physical Conditioning and SimPPW Military Training tended to view themselves as more capable of performing work requiring high levels of Energy compared to other types of work. Conversely, Soldiers who score high on ASVAB GT and the LeadEx tended to view themselves as less capable of performing work requiring high levels of Energy compared to other types of work.

Table 5.4. Intercorrelations between WSI and Original Leadership Assessment Tool Scales for E5 Soldiers

Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
EMF: ASVAB GT Score	**-.10**	-.04	-.02	**-.14**	**-.20**	.02	**-.13**	**.17**	.00	**.11**	.06	.08	.09	.05	.09	-.02
LeadEx: 24-Item Composite	.01	.01	.06	-.03	**-.15**	-.02	**-.14**	.08	.04	.01	**.10**	-.07	-.01	**.14**	.01	-.03
LeadEx: 40-Item Composite	.01	.04	.07	-.02	**-.15**	-.05	**-.14**	.09	.04	.00	**.10**	-.06	-.04	**.11**	-.01	-.01
Simulated PPW Composite	.09	-.04	.07	-.02	**-.10**	-.08	.03	.05	.03	.01	.00	-.01	.03	.02	-.07	-.01
PPW Awards	-.01	.02	**.11**	.00	-.04	-.04	.01	.06	-.03	.08	-.08	.09	-.08	-.01	-.07	-.02
PPW Military Education	.08	.03	.01	.04	-.05	-.05	.05	-.02	.00	.00	-.01	-.07	.04	.07	**-.10**	.00
PPW Civilian Education	.08	-.06	.02	.00	-.01	-.04	**-.11**	.09	.07	-.03	-.03	.02	.01	.00	-.05	.02
PPW Military Training	.00	-.09	.02	**-.12**	**-.15**	-.05	**.17**	-.01	.01	.00	**.15**	-.04	.08	-.03	**.13**	-.05
ExAct: Computer Experience	.00	-.04	.02	-.07	-.05	-.01	.05	.05	.09	**.16**	.01	.02	-.06	-.07	-.01	-.07
ExAct: Supervisory Experience	**.20**	-.04	.05	-.07	-.07	.05	.08	-.03	.06	.00	**.13**	-.01	**-.09**	-.06	.03	**-.18**
ExAct: General Experience	**.13**	-.02	.04	**-.21**	**-.10**	-.04	.08	.00	.07	.06	**.11**	.01	.02	**-.13**	**.12**	**-.10**
SDI: Dependability	.06	-.08	.00	.08	-.01	-.07	-.07	.04	.01	.03	.03	-.02	-.02	.00	-.03	.02
SDI: Adjustment	-.04	-.04	**-.10**	**-.13**	**-.10**	**-.12**	-.07	.04	.01	**.16**	.06	-.03	**.15**	.03	**.14**	.04
SDI: Work Orientation	**.23**	.00	**.10**	**-.12**	**-.19**	.04	**.10**	**-.11**	**.14**	.05	**.12**	.01	-.06	**-.12**	.04	**-.16**
SDI: Agreeableness	.03	-.01	-.04	.08	.03	-.07	-.09	.00	-.02	.07	-.08	-.04	**.10**	.02	.03	-.04
SDI: Physical Conditioning	.05	-.03	-.01	.01	-.03	-.05	**.18**	-.09	-.06	.07	.07	**-.12**	.03	.00	.03	-.05
SDI: Leadership	.07	-.05	.03	**-.19**	**-.21**	-.06	-.04	-.01	.08	**.13**	**.24**	.05	-.06	.03	**.14**	**-.10**
IQ-II: Hostility to Authority	**-.17**	-.04	-.08	.06	.02	.07	.06	-.05	.03	**-.12**	-.02	.01	.05	.02	.04	.09
IQ-II: Manipulativeness	**-.12**	**-.11**	-.04	.03	.06	**.12**	**.12**	.01	-.01	-.05	**-.09**	.02	.04	.01	-.08	.08
IQ-II: Social Perceptiveness	.00	.00	-.03	-.05	-.03	-.08	.04	**-.10**	.00	**.11**	.05	-.05	-.06	.01	**.11**	.09
IQ-II: Tolerance for Ambiguity	-.02	.01	-.05	**-.22**	**-.15**	-.07	-.06	-.09	**.18**	**.18**	**.14**	.01	-.05	.04	**.23**	-.04
IQ-II: Emergent Leadership	.05	**-.17**	-.03	**-.13**	**-.11**	-.08	.05	-.05	.05	**.15**	**.24**	.02	-.02	-.03	**.17**	-.06
IQ-II: Interpersonal Skills	-.05	-.02	-.05	.01	-.03	**-.15**	**-.11**	-.01	.04	**.15**	.01	-.01	.03	**.10**	**.11**	.00

Note. n = 312-322. Statistically significant correlations are bolded, $p < .05$ (one-tailed). Each column corresponds to a different WSI scale: 1 = Achievement/Effort; 2 = Adaptability/Flexibility; 3 = Attention to Detail; 4 = Concern for Others; 5 = Cooperation; 6 = Dependability; 7 = Energy; 8 = Independence; 9 = Initiative; 10 = Innovation; 11 = Leadership Orientation; 12 = Persistence; 13 = Self-Control; 14 = Social Orientation; 15 = Stress Tolerance; 16 = Cultural Tolerance.

Soldiers who scored high on ASVAB GT tended to view themselves as more capable of performing work requiring Independence and Innovation compared to other types of work. Additionally, Soldiers who scored high on SDI Adjustment, IQ-II Social Perceptiveness, and IQ-II Tolerance for Ambiguity tended to view themselves as more capable of performing work requiring Innovation compared to other types of work. Such findings suggest that intelligent and cognitively flexible Soldiers feel they would be best at types of work that require them to work independently and deal with tasks that require novel solutions.

Not surprisingly, Soldiers who scored high on the SDI and IQ-II Leadership scales tended to view themselves as more capable of performing work requiring Leadership Orientation compared to other types of work. Though the relations were not as strong as they were for SDI and IQ-II Leadership, Soldiers who scored high on the ExAct Supervisory and General Experience scales, SDI Work Orientation, and IQ-II Tolerance for Ambiguity tended to view themselves as more capable of performing work requiring Leadership Orientation compared to other types of work. Soldiers who scored high on IQ-II Tolerance for Ambiguity, Emergent Leadership, and Social Perceptiveness tended to view themselves as more capable of performing work requiring Stress Tolerance compared to other types of work.

Lastly, relations between the original LAT instrument scales and the remainder of the WSI scores not mentioned above (e.g., Adaptability/Flexibility, Dependability, Initiative, Persistence, Social Orientation, and Cultural Tolerance) tended to be inconsistent across pay grades. Such inconsistencies may stem from Soldiers differential interpretation of the statements used to represent these traits on the WSI (recall, the WSI uses a single statement to broadly define each trait).

Predictor and Soldier Website Versions of SimPPW and ExAct

Table 5.5 shows the correlations between scales on the predictor and Soldier website versions of the SimPPW and ExAct scales. Because of the relatively small within-pay-grade sample sizes, we focus on the combined sample results here. Similar to the predictor versions of these instruments for E4 Soldiers, the SimPPW Composite had larger correlations with ExAct Supervisor and General Experience than with ExAct Computer Experience. This relatively low correlation with ExAct Computer Experience extended itself to the other SimPPW scale scores except SimPPW Civilian Education. The pattern of correlations was similar to that of the predictor versions of these instruments for E4 Soldiers, but not for E5 Soldiers.

Table 5.5. Intercorrelations between Predictor and Soldier Website Versions of SimPPW and ExAct Scales

Predictor Version Scale	Soldier Website Version Scale							
	1	2	3	4	5	6	7	8
E4 Soldiers								
SimPPW								
1. SimPPW Composite	**.43**	**.36**	**.28**	**.31**	.16	**.24**	**.25**	**.37**
2. Awards	.21	**.57**	.07	-.07	.11	.04	.23	**.47**
3. Military Education	**.23**	.02	**.32**	.09	.00	.09	.16	.04
4. Civilian Education	.15	-.04	.04	**.51**	**-.21**	.15	-.18	-.20
5. Military Training	**.44**	**.24**	**.28**	**.23**	**.52**	**.30**	**.39**	**.53**
ExAct								
6. Computer Experience	-.10	-.17	-.07	.10	**-.20**	**.44**	-.01	.07
7. Supervisor Experience	**.28**	.15	**.32**	-.03	**.24**	**.24**	**.53**	**.53**
8. General Experience	**.25**	**.27**	.19	-.03	**.30**	**.29**	**.43**	**.73**
E5 Soldiers								
SimPPW								
1. SimPPW Composite	**.52**	**.26**	**.38**	**.50**	.19	**.22**	.10	**.24**
2. Awards	**.26**	**.40**	**.30**	.05	-.09	.18	.06	.17
3. Military Education	**.25**	.15	**.26**	.13	.04	.08	-.01	.11
4. Civilian Education	**.48**	.07	**.25**	**.74**	.10	**.23**	.09	.12
5. Military Training	.03	.05	-.03	-.15	**.43**	-.06	.12	.17
ExAct								
6. Computer Experience	**.38**	.07	**.33**	**.35**	.10	**.42**	-.10	.17
7. Supervisor Experience	**.21**	.13	.07	**.26**	.19	.09	**.50**	**.34**
8. General Experience	**.40**	**.26**	**.27**	**.27**	**.35**	**.28**	**.32**	**.52**
E4 and E5 Soldiers Combined								
SimPPW								
1. SimPPW Composite	**.59**	**.50**	**.39**	**.47**	**.32**	**.18**	**.41**	**.55**
2. Awards	**.41**	**.64**	**.26**	**.15**	**.21**	.08	**.41**	**.60**
3. Military Education	**.36**	**.25**	**.33**	**.20**	**.14**	.07	**.24**	**.29**
4. Civilian Education	**.44**	**.17**	**.23**	**.68**	.09	**.19**	.12	**.18**
5. Military Training	**.28**	**.23**	**.16**	.08	**.50**	.12	**.34**	**.42**
ExAct								
6. Computer Experience	**.20**	.03	**.17**	**.26**	.01	**.42**	.05	**.20**
7. Supervisor Experience	**.42**	**.39**	**.29**	**.22**	**.34**	.13	**.65**	**.67**
8. General Experience	**.47**	**.47**	**.30**	**.24**	**.42**	**.22**	**.56**	**.78**

Note. n_{E4} = 70-73. n_{E5} = 65-68. $n_{Combined}$ = 135-141. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

Longitudinal Validity Analyses

Table 5.6 presents raw and corrected validity estimates for each predictor score. Few of the raw correlations were significant relative to the concurrent validation results. This was at least partly due to the low power associated with the small sample sizes ($n_{Observed}$ = 55-56; $n_{Expected\ Future}$ = 52-53). In this context, the results in this table should be interpreted with some

caution. Some observations are notable, however. The corrected validity estimates for the SimPPW composite are similar to those from the concurrent validation effort for E5s (.19 and .13 for the Observed and Expected Future Composites, respectively). In addition, SDI Work Orientation was the predictor with the highest corrected correlation with the Observed and Expected Future Composites in the concurrent validation for E5 Soldiers. In this longitudinal validation, Work Orientation had the highest corrected correlation with the Observed Performance Composite and a relatively high correlation with the Expected Future Performance Composite.

Beyond these observations, there were a number of correlations above .20, but the sample sizes were not sufficient for many of these to be significant. Another consequence of the small sample sizes is that examining the incremental validity of the predictors over the PPW composite was not practical.

Table 5.6. Raw and Corrected Correlations between Predictor and Ratings Criterion Scores

	Observed Performance Composite		Expected Future Performance Composite	
Predictor	Raw	Corrected	Raw	Corrected
SimPPW Composite	.12	.17	.10	.18
ASVAB GT	-.04	-.05	.07	.12
LeadEx24	.18	.26	**.26**	.44
LeadEx40	.12	.17	.23	.39
ExAct: Computer	-.13	-.19	-.18	-.31
ExAct: Supervisory	.01	.01	-.08	-.14
ExAct: General	.06	.08	.05	.08
SDI: Dependability	.11	.16	.12	.20
SDI: Adjustment	-.02	-.02	.21	.36
SDI: Work Orientation	**.27**	.38	**.24**	.42
SDI: Leadership	.16	.23	.20	.34
SDI: Agreeableness	.05	.08	**.33**	.56
SDI: Physical Conditioning	-.02	-.03	.09	.16
IQ-II: Tolerance of Ambiguity	-.14	-.20	-.17	-.30
IQ-II: Interpersonal Skills	-.02	-.03	**.24**	.41
IQ-II: Social Perceptiveness	**-.30**	-.43	-.19	-.33
IQ-II: Emergent Leadership	.04	.06	.03	.06
IQ-II: Manipulativeness	-.02	-.02	-.16	-.28
IQ-II: Hostility	-.02	-.03	-.18	-.32
WSI: Achievement/Effort	-.02	-.02	-.17	-.28
WSI: Adaptability/Flexibility	-.02	-.03	.17	.29
WSI: Attention to Detail	.05	.07	**-.27**	-.46
WSI: Concern for Others	-.09	-.12	-.07	-.13

Table 5.6. (Continued)

Predictor	Observed Performance Composite		Expected Future Performance Composite	
	Raw	Corrected	Raw	Corrected
WSI: Cooperation	-.08	-.12	-.07	-.12
WSI: Dependability	.21	.31	**.28**	.47
WSI: Energy	.10	.14	-.17	-.29
WSI: Independence	-.08	-.11	-.19	-.32
WSI: Initiative	-.01	-.01	-.09	-.15
WSI: Innovation	-.20	-.29	-.06	-.11
WSI: Leadership Orientation	-.04	-.06	.04	.06
WSI: Persistence	.09	.12	.03	.05
WSI: Self-Control	.06	.08	.12	.20
WSI: Social Orientation	.05	.08	.22	.38
WSI: Stress Tolerance	-.12	-.17	.03	.05
WSI: Cultural Tolerance	.09	.14	.17	.29

Note. n_{Observed} = 55-56; $n_{\text{Expected Future}}$ = 52-53. "Corrected" correlations were corrected for criterion unreliability. Interrater reliability estimates for E5 Soldiers from the NCO concurrent validation effort were used for this correction (Observed Performance Composite single rater reliability for E5 Soldiers = .45; Expected Future Composite single rater reliability for E5 Soldiers = .31). Statistically significant correlations are bolded, $p < .05$ (one-tailed).

As discussed in Chapter 2, promotion during the research period was developed as an alternative criterion in response to the low number of Soldiers with performance ratings. Table 5.7 shows the results of an analysis in which each predictor and exposure were used to predict promotion. Fortunately, the sample sizes for these analyses were more substantial (n_{E4} = 469-510. n_{E5} = 242-259).

As discussed in Chapter 2, exposure reflects the number of months a Soldier had been eligible to be promoted at the time the data collection ended or the Soldier left the Army, whichever came first. Also recall that the exposure variable was developed for two reasons. First, we wished to ensure that only those Soldiers who had some minimal opportunity to be promoted in terms of exposure were included in validity analyses. The value of 6 months was selected as a reasonable minimal period. Second, exposure itself could be a predictor of promotion or affect the relation between the other predictors and the promotion criterion.

The values in the Step 1 columns of Table 5.7 are point-biserial correlations between each predictor and promotion. Next, using logistic regression, Exposure and each predictor were used to estimate the probability of promotion. The values in the Step 2 columns are the correlations between these predicted probabilities and actual promotion. Step 1 reflects the extent to which each predictor scale was predictive of promotion on its own. Step 2 reflects the extent to which each predictor scale and exposure, together, were predictive of promotion. Bolded values indicate the correlations that were significantly different from zero. The superscripted "a" indicates that hierarchical logistic regression determined that exposure significantly incremented the prediction of promotion beyond the predictor. It is important to note that the promotion base rates for E4 and E5 Soldiers were 37.3% and 27.4%, respectively. As the base rate for a dichotomous variable varies from .50 the potential range of the point-biserial correlation is restricted.

Table 5.7. Correlations of Predictors and Exposure with Promotion Criterion

	E4 Soldiers		E5 Soldiers	
	Step 1	Step 2	Step 1	Step 2
Exposure	**.13**		-.04	
SimPPW Composite	**.25**	**.26**	**.21**	**.23**
ASVAB GT Score	-.07	**.17** [a]	.10	.12
LeadEx_24	**.09**	**.15** [a]	-.04	.05
LeadEx_40	.08	**.14** [a]	-.05	.06
ExAct: Computer	.06	**.13** [a]	-.04	.05
ExAct: Supervisory	**.28**	**.28**	**.18**	**.20**
ExAct: General	**.21**	**.21**	.10	.11
SDI: Dependability	.07	**.15** [a]	.02	.04
SDI: Adjustment	.06	**.14** [a]	.07	.07
SDI: Work Orientation	**.15**	**.20** [a]	.13	**.14**
SDI: Leadership	**.11**	**.17** [a]	**.17**	**.17**
SDI: Agreeableness	.02	**.13** [a]	-.04	.04
SDI: Physical Conditioning	**.10**	**.17** [a]	.08	.07
IQ-II: Tolerance of Ambiguity	.01	**.12** [a]	.06	.07
IQ-II: Interpersonal Skills	**.10**	**.16** [a]	.00	.04
IQ-II: Social Perceptiveness	-.01	**.12** [a]	.06	.07
IQ-II: Emergent Leadership	**.13**	**.18** [a]	**.15**	**.15**
IQ-II: Manipulativeness	**-.12**	**.16** [a]	-.11	**.13**
IQ-II: Hostility	-.09	**.15** [a]	-.02	.05
WSI: Achievement/Effort	.12	**.18** [a]	.08	.10
WSI: Adaptability/Flexibility	-.02	**.14** [a]	-.04	.06
WSI: Attention to Detail	.06	**.14** [a]	-.02	.04
WSI: Concern for Others	-.02	**.13** [a]	.02	.05
WSI: Cooperation	-.07	**.15** [a]	.00	.04
WSI: Dependability	.08	**.16** [a]	-.03	.05
WSI: Energy	.01	**.14** [a]	.08	.09
WSI: Independence	-.06	**.15** [a]	-.11	.12
WSI: Initiative	.09	**.16** [a]	.07	.08
WSI: Innovation	.00	**.13** [a]	.06	.07
WSI: Leadership Orientation	.06	**.14** [a]	.08	.09
WSI: Persistence	-.05	**.14** [a]	.07	.08
WSI: Self-Control	-.03	**.14** [a]	-.03	.05
WSI: Social Orientation	-.04	**.14** [a]	-.04	.05
WSI: Stress Tolerance	-.04	**.14** [a]	.07	.08
WSI: Cultural Tolerance	-.07	**.15** [a]	-.11	.11

Note. n_{E4} = 469-510. n_{E5} = 242-259. Step 1 = the raw correlation between predictor and promotion. Step 2 = the raw correlation between the (a) predicted probability of promotion based on the predictor and exposure and (b) promotion. Statistically significant correlations are bolded, $p < .05$ (one-tailed).

[a] The increment in prediction between Step 1 and Step 2 is statistically significant, $p < .05$.

Table 5.7 contains a number of interesting results. First, adding exposure significantly incremented the prediction of promotion for many of the predictors in the E4 Soldier sample but for none of the predictors in the E5 Soldier sample. Second, the correlations between the

SimPPW composite and promotion are relatively high for both pay grades. The main reason for this is that actual PPW scores are a primary operational determinant of promotion. There are a number of likely reasons why the correlations between SimPPW score and promotion were not even higher, including the following:

- Promotion rates and the level of PPW scores required to get promoted vary considerably across MOS.
- Soldiers do not become eligible for promotion unless their commander recommends them.
- The SimPPW scores do not include the points that come with the commander's recommendation or the promotion board, but there tends to be very little variation in these.
- Finally, the Soldiers' actual PPW points were very likely not the same when they completed the PFF21 as part of the predictor data collection compared to when their operational PPW was forwarded for promotion.

The conceptual relation that SimPPW and operational PPW scores have with promotion is so direct that examining incremental validity estimates of the predictors beyond SimPPW would not be a good indication of the ability of the experimental predictors to predict performance beyond the PPW.

Three other sets of analyses were explored and are discussed here briefly. First, the hypothesis that the relation between exposure and promotion could be non-linear was examined. For example, up to a certain number of months of exposure, the probability of promotion might go up, after which additional exposure relates negatively to the probability of promotion. This hypothesis was not supported by the analyses. Second, the semi-partial correlations between the predictors and promotion, with the variance due to exposure removed from promotion, were compared to the raw correlations between the predictors and promotion to determine whether exposure could be masking relations between the predictors and promotion. The semi-partial correlations were not larger than the raw correlations. Finally, there was an attempt to identify MOS with sufficient sample sizes to calculate within-MOS point-biserial correlations between the predictors and promotion. After eliminating Soldiers without the data to support the calculation of exposure and Soldiers who did not have at least 6 months of exposure, no MOS had sufficient sample sizes to support these analyses.

Despite the (a) incomplete overlap in content between the SimPPW Composite and the operational promotion criterion and (b) deficiency (i.e., job performance possibly includes elements beyond those assessed by the operational promotion criterion) and contamination (e.g., MOS membership also effects the probability of promotion) in the promotion criterion, Table 5.7 shows some interesting results for the predictors. ExAct Supervisory Experience had a relatively strong correlation with promotion for E4 and E5 Soldiers. ExAct General Experience performed similarly, but only for E4 Soldiers. These values are larger than those observed for E5s in the concurrent validity analysis when performance rating composites were the criterion. It is noteworthy that, in the concurrent validity results, ExAct scores showed no incremental validity beyond the SimPPW composite and in these results exposure does not increase the validity

estimates for ExAct Supervisory and General Experience. This pattern of correlations and the content of the PFF21, from which SimPPW scores were derived, provide support for the hypothesis that the ExAct Supervisory and General Experience scores are construct-valid measures of experience. The validity estimate for the 24-Item LeadEx suggests that this situational judgment test is somewhat predictive of promotion. LeadEx results were more favorable in the concurrent validity results when performance ratings were the criterion. The two temperament instruments (i.e., SDI and IQ-II) also showed some positive results in these analyses. SDI Leadership performed relatively well compared to the other scales for E4 and E5 Soldiers. SDI Work Orientation and Physical Conditioning did relatively well for E4 Soldiers. SDI Leadership, Work Orientation, and Physical Conditioning were among the stronger scales in the concurrent validation as well. The IQ-II also showed somewhat positive results in this research. IQ-II Emergent Leadership performed relatively well for both E4 and E5 Soldiers and IQ-II Interpersonal Skills and Manipulativeness showed some positive results for E4s. These three scales were among four that showed significant correlations for this instrument with E5 Soldier observed performance ratings in the concurrent validation sample. The WSI scores did not show any significant correlations with promotion; however, the WSI was not one of the original LAT predictor measures. It was designed for predicting the performance of first-term Soldiers. In another effort, it showed good results predicting first-term performance using other methods of scoring (McCloy & Putka, 2006).

Summary/Discussion

Taken together with the results presented in Chapter 3, the results for the predictor versions of the NCO21 predictor measures indicate that their functioning remained largely the same among Soldiers in this longitudinal validation data collection relative to Soldiers in the concurrent validation sample. The pattern of correlations among the scale scores for the instruments administered during the predictor portion of this project were remarkably similar to those observed in the concurrent validation effort. Again, these similarities were present despite the fact that the mode of administration for the instruments differed across samples.

The overall sample (i.e., E4 and E5 Soldiers Combined) across-instrument correlations between the predictor data collection and Soldier Website versions of the PPW and ExAct scores were similar to the correlations among these scales in the E4 Soldier predictor sample but not the E5 Soldier sample. Given the necessity to combine the E4 and E5 Soldier samples for this analysis, this finding is reasonable evidence in support of the construct validity of these measures. Their construct validity was well supported by the original concurrent validation evidence.

The primary goal of this research was to collect evidence regarding the longitudinal criterion-related validity of the NCO21 predictor measures referred to as the LAT. The major difficulty was a low response rate on the NCO Promotion Soldier website that resulted in usable performance ratings for only 56 of the 942 Soldiers who completed the LAT during the predictor portion of this project. The sizes of the longitudinal criterion-related validity estimates (i.e., correlations between predictor scale scores and job performance ratings) were encouraging regarding the predictive capacity of these instruments. However, the sample size was so small

that very few of the correlations were significant, thus weakening potential inferences regarding validity.

In the context of this difficulty, promotion during the research project period was used as an alternative criterion. This criterion's close operational relation with the PPW is a substantial caveat. Nevertheless, scales from each of the original NCO21 predictors (i.e., LeadEx, ExAct, SDI, and IQ-II) showed positive longitudinal validity results that were fairly consistent with the concurrent validation results.

CHAPTER 6: SUMMARY

This research had three primary goals. The first was to examine whether the evidence supporting the concurrent criterion-related validity of the experimental predictors in the LAT would extend to the longitudinal validation setting. Another goal of this research was to examine the extent to which it would be practical and psychometrically reasonable to collect data on the predictor measures via laptop computer instead of paper-and-pencil. The third goal was to determine whether it would be practical and psychometrically reasonable to collect criterion data (i.e., job performance ratings) via e-mail and the Internet instead of paper-and-pencil in a standardized data collection setting.

Empirical Results for Longitudinal Criterion-Related Validity

The longitudinal validity results using performance ratings as a criterion were promising in terms of the size of the validity estimates, but the small sample size yielded too little power and thus few of the estimates were statistically significant. In addition, the sample was sufficiently small that we had too few Soldiers to support examining the (a) incremental validity of the LAT predictors beyond the SimPPW and (b) validities by pay grade. Results were more promising when promotion was the criterion. Sample size was not a problem for separate E4 and E5 Soldier analyses. However, as described in Chapter 5, the relation between SimPPW scores, operational PPW scores, and the promotion criterion mean these validity estimates need to be interpreted with some caution, primarily because the operational PPW is a substantial contributor to the promotion decision. Just the same, scales from each of the experimental predictors showed positive longitudinal validity results that were fairly consistent with the concurrent validation results.

Collecting Data on the Computer

Together with the original NCO21 concurrent validation effort, three methods of data collection were employed. The concurrent data collection almost exclusively used a paper-and-pencil approach, with the instruments administered during monitored testing sessions (Knapp et al., 2004). The predictor data for this longitudinal validation research were collected during monitored sessions using testing software on laptop computers. The primary criterion data for this longitudinal research were collected via the Internet. Accordingly, this research project afforded a good opportunity to examine the transition from paper to computer-based methods of data collection because the same instruments were used in both the concurrent (i.e., paper based) and longitudinal (i.e., computer based) validity data collections.

Instrument and System Development

Early on, we needed to select software suitable for administering the predictor and criterion instruments. Questionmark's Perception® package was selected for the laptop computer administration of predictors, and the PERL programming language was selected for development of the NCO Promotion Soldier and Supervisor websites for criterion data collection.

Next, decisions were made regarding general characteristics of the tools that illustrate some of the advantages of computerizing the instruments. For example, in the concurrent

validation effort, none of the measures had firm time limits. This characteristic, in combination with the computer's capacity to record each individual's completion time for each measure, gave us the opportunity to collect accurate information about how long each measure takes. Another decision was whether to present items or rating scales one at a time or to present one screen's worth of items at a time. If participants did not respond, they were warned and given the opportunity to answer the item or move on to the next item. This feature was designed to reduce the incidence of "missing data." Another advantage of these computerized measures was that the item stem and response options were presented together on the same screen on which the Soldier responded. This format afforded the participants a smoother and less complicated experience relative to the concurrent validation effort during which Soldiers had to deal with test booklets and scannable answer sheets. In addition, computer administration allowed for the electronic collection and processing of data, thus eliminating some of the errors associated with processing scannable sheets. Also, the administration software monitored out-of-range and illogical responses and gave the participant an opportunity to correct them. For example, on the predictor data collection and the criterion data collection (i.e., NCO Promotion Soldier website) versions of the PFF21, if the Soldiers indicated that they had passed the APFT, they were asked to enter their score. If the Soldiers had passed the APFT, the score ranges from 180 to 300. If Soldiers entered an out-of-range response, they got a warning and were asked to revisit the question. All of these characteristics of the laptop computer and Internet methods of administration provide advantages over paper-and-pencil administration.

Nevertheless, there were some challenges associated with the logistics of managing computerized data collection. The predictor data collection portion of the research project involved traveling to a number of U.S. Army installations to administer the measures. The general procedure was to administer the predictor measures to groups of up to 30 Soldiers at a time on IBM Notebook computers on which we had previously installed Perception software containing the LAT. The computers were shipped to and from the installations in customized carrying cases. During the development of the laptop computerized version of the predictors, the obvious advantage of less paper became apparent (e.g., no preparation, management, and transportation of "Soldier packets" containing test booklets and scannable answer sheets). However, it is important to keep in mind that these activities were replaced by (a) loading the LAT and its software on many computers, (b) shipping large cases that needed secure storage to points-of-contact, and (c) setting up computers in rooms that were not always optimal for this type of activity. This took considerable time and effort for which planning was required.

The development of the NCO Promotion Soldier and Supervisor websites for the criterion portion of the data collection was somewhat more straightforward. This was largely because the PERL programming language, despite requiring more "from scratch" programming, required fewer formatting compromises than the Perception software. However, an extensive amount of work was done to develop the system and materials to send solicitation e-mails, participation e-mails, and reminders to Soldiers and Supervisors. Monitoring and processing responses took time as well.

How Well Did Computer Data Collection Work?

Collecting the data via laptop computers worked well after some development time. Data collection sessions were efficient, and Soldiers liked using computers instead of managing item

booklets and scannable answer sheets. As the results presented in Chapter 3 show, the psychometric characteristics of the instruments were robust to the transition from paper-and-pencil to laptop computer administration. Table 6.1 compares paper and laptop administration data collection efficiency. The "% Missing Data" column shows the percentage of Soldiers whose data were dropped for a particular instrument because they responded to fewer than 90% of the items. For all but the ExAct, laptop administration resulted in less data loss on this index than did paper-and-pencil administration. The" % Testing Time" column shows the percentage of Soldiers whose data were dropped for a particular instrument because they completed the instrument in an unreasonably short time. It is not practical to collect this information during group paper-and-pencil administrations. Computer administration allows the researcher to identify individuals who complete an instrument too quickly and remove their data from the data sets, thus providing more accurate data. Finally, the "% Response Pattern" column shows the percentage of Soldiers whose data were dropped because they exhibited pattern responding (e.g., given option A, B, C or D, using option A too frequently). Laptop administration of the LeadEx, SDI, and IQ-II resulted in less data loss for this reason compared to the paper-and-pencil administration of these instruments. Taken together, together the results suggest that administering the LAT predictor instruments via laptop computers is efficient and produces high-quality data.

Table 6.1. Comparison of Administration Methods in Terms of Data Collection Efficiency

	Reason for Data Loss								
	% Missing Data			% Testing Time			% Response Pattern		
Instrument	Paper	Laptop	Internet	Paper	Laptop	Web	Paper	Laptop	Internet
PFF21	0.11	0.00	2.84	N/A	0.00	0.00	0.00	0.00	0.00
ExAct	0.58	0.85	3.55	N/A	1.38	0.00	0.00	0.00	0.71
LeadEx	1.64	0.53	N/A	0.79[a]	4.99	N/A	0.26	0.00	N/A
SDI	1.97	0.42	N/A	N/A	0.74	N/A	0.58	0.42	N/A
IQ-II	1.96	0.53	N/A	N/A	0.64	N/A	1.39	0.74	N/A

Note. n_{Paper} = 1877-1891. n_{Laptop} = 942. n_{Web} = 141. % Missing data = Percentage of Soldiers who failed to respond to at least 90% of the instrument's items. % Testing Time = Percentage of Soldiers who completed the instrument in an unreasonably short time. % Response Pattern = Percentage of Soldiers who exhibited patterned responding on the instrument. Paper = Data from paper-and-pencil administration of instruments during the NCO21 concurrent validation effort (Knapp et al., 2004). Laptop = Data from laptop computer administration of instruments during this project's predictor data collection. Web = Data from Internet administration on the NCO Promotion Soldier website during this project's criterion data collection.

[a] These Soldiers (n = 15) were eliminated from further analyses because they did not finish the test, meaning that the statistics for the last few items might have been distorted.

Collecting the data remotely via the Internet, however, did not work as well as paper-and-pencil or laptop computer administration in a proctored test setting. Comparisons of subgroup differences and correlations among scales, within and across versions, suggest that the PFF21 and ExAct (i.e., the LeadEx, SDI, and IQ-II were not administered via the Internet) functioned similarly across time and modes of administration. However, Table 6.1 shows a relatively higher data loss rate for the PFF21 and the ExAct due to missing data and a higher loss rate for pattern responding for the ExAct. Similar results were observed when comparing the paper-and-pencil based collection to the Internet-based collection of supervisor job performance ratings. Loss of data due to missing data and pattern responding was 7.2% and 2.6% for observed and expected future performance ratings, respectively, for the paper-and-pencil administration during the

concurrent validation effort. The comparable loss was 12.0% and 18.8%, respectively, for the NCO Promotion Supervisor Website. These results were likely due to the absence of a standardized administration situation with administrators present to monitor and motivate Soldier and supervisor participants.

Level of Participation

The big problem with collecting data remotely via e-mail and the Internet was the level of participation. As shown in Chapter 2, participation rates were very low in terms of Soldiers who responded to the solicitation and participation e-mails by logging on to the NCO Promotion Soldier website. A major difference between the paper-and-pencil and laptop computer data collections compared to the e-mail and Internet approach was whether data collectors went to the site or recruited participants via e-mail. Participation levels were much higher when we were on-site with the Soldiers versus recruiting via e-mail. However, this relatively low level of participation based on e-mail recruitment is confounded by this research project being a longitudinal data collection. Part of the issue is recruiting the participation of a particular Soldier, who participated in first phase of the project, instead of any Soldier who meets some set of demographic requirements (e.g., pay grade and/or TIS). Therefore, in this context, there are two problems to manage or solve: (a) finding a particular Soldier, and (b) recruiting that Soldier remotely.

Collecting Soldier e-mail addresses during the predictor data collection was a fairly effective method of identifying Soldiers for the criterion data collection. The solicitation and participation e-mails were sent to 926 of the original 942 Soldiers who participated in the predictor data collection. The problem was that only 141 of these Soldiers logged on to the NCO Promotion Soldier website and agreed to participate. There was some loss of participation when all of the solicited supervisor raters did not respond, but this problem was not nearly as serious.

Lessons Learned

In longitudinal data collections, it is not unusual to have problems finding and recruiting participants after the first data collection. For longitudinal validation efforts to work, however, this problem needs to be solved or at least managed. One approach would be to include a much larger number of participants during the first data collection so that the loss across data collection events does not drive the sample size below a predetermined cutoff. Depending on the circumstances, however, this solution could be prohibitively expensive. Another strategy would be for the participation solicitation e-mail to come from a very important person (VIP) who is organizationally more proximate to the Soldier being recruited (e.g., a division or installation level commander). This method would require more coordination but might make the Soldier more likely to respond. In addition, it might help to communicate with the participating Soldiers between the predictor and criterion data collections. For example, a project newsletter could be sent to the participants via e-mail each month that could (a) ask them to update contact information, (b) remind them how important the project is, and (c) tell them when the next request for participation will be coming.

Another potential approach would be to identify criteria that the Army already collects as part of its administrative records. The alternate criterion in this project (promotion) is an

example. Gathering data for elements of the operational PPW would be another example. In this project, our attempt to gather these types of archival data in cooperation with the Army Human Resources Command (AHRC) did not result in sufficient data for useful analysis. As records are increasingly computerized and human resource systems integrated, however, such an approach might become more practical. A significant caution to keep in mind with archival data is that they rarely measure all the conceptual parts of the job performance space that researchers are interested in. For example, job performance ratings can address almost any dimension of interest. Archival data tend to have a narrower focus. However, if archival data could be meaningfully used as criteria, it could substantially address the participation problem.

A Final Word

This research developed some evidence supporting the longitudinal validity of the LAT predictor measures. Additional research in a more operational setting is recommended, however, to support the assignment of promotion points in the Army's semi-centralized NCO promotion system based on any of these measures. This research also showed that collecting data using laptop computers is psychometrically reasonable and probably more efficient than paper-and-pencil data collection. Data collection via e-mail and the Internet, however, was not particularly effective at ensuring sufficient rates of participation. Strategies for addressing this issue were discussed in this chapter.

References

Borman, W.C., Kubisiak, U.C., & Schneider, R.J. (1999). Work styles. In N.G. Peterson, M.D. Mumford, W.C. Borman, P.R. Jeanneret, & E.A. Fleishman (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 213-226). Washington D.C.: American Psychological Association.

Department of the Army. (2004). *Army Regulation 600-8-19: Enlisted Promotions and Reductions*. Washington, D.C.: Department of the Army.

Ford, L.A., Campbell, R.C., Campbell, J.P., Knapp, D.J., & Walker, C. B. (2000). *21st Century Soldiers and Noncommissioned Officers: Critical Predictors of Performance* (Technical Report 1102). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hicks, L.E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*, 167-184.

Kilcullen, R.N., Chen, G., Zazanis, M. M., Carpenter, T., & Goodwin, G. (1999, April). *Adaptable performance in unstructured environments.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Kilcullen, R.N., Mael, F.A., Goodwin, G.F., & Zazanis, M.M. (1999). Predicting U.S. Army Special Forces Field Performance. *Journal of Human Performance in Extreme Environments, 4*, 53-63.

Kilcullen, R.N., White, L.A., Zacarro, S., & Parker, C. (2000, April). *Predicting managerial and executive performance.* Paper presented at the 15th annual meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Knapp, D.J., Burnfield, J.L, Sager, C.E., Waugh, G.W., Campbell, J.P., Reeve, C.L., Campbell, R.C., White, L.A., & Heffner, T.S. (2002). *Development of predictor and criterion measures for the NCO21 research program* (Technical Report 1128). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., McCloy, R.A., & Heffner, T.S (Eds.) (2004). *Validation of Measures Designed to Maximize 21st-Century Army NCO Performance* (Technical Report 1145). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., Sager, C.E., & Tremble, T.R. (Eds.) (2005). *Development of Experimental Army Enlisted Personnel Selection and Classification Tests and Job Performance Criteria* (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McCloy R.A., & Putka, D.J. (2005). The work suitability inventory. In D. Knapp, C. Sager, & T. Tremble (Eds.), *Development of Experimental Army Enlisted Personnel Selection and Classification Tests and Job Performance Criteria* (Interim Technical Report IR-05-05). Alexandria, VA: Human Resources Research Organization.

McCloy R.A., & Putka, D.J. (2006). Work suitability inventory. In D. Knapp & T. Tremble (Eds.), *Concurrent Validation of Experimental Army Enlisted Personnel Selection and Classification Measures* (HumRRO Final Report FR-06-35). Alexandria, VA: Human Resources Research Organization.

Putka, D.J. (2004). Experiences and Activities Records (ExAct). In D. Knapp, R. McCloy, & T. Heffner (Eds.), *Validation of Measures Designed to Maximize 21st-Century Army NCO Performance* (Technical Report 1145). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Putka, D.J., & Campbell, R. C. (2004). Simulated promotion point worksheet (SimPPW). In D. Knapp, R. McCloy, & T. Heffner (Eds.), *Validation of Measures Designed to Maximize 21st-Century Army NCO Performance* (Technical Report 1145). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Putka, D.J., Kilcullen, R.N., & White, L.A. (2004). Temperament inventories. In D. Knapp, R. McCloy, & T. Heffner (Eds.), *Validation of Measures Designed to Maximize 21st-Century Army NCO Performance* (Technical Report 1145). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Sager, C.E., Putka, D.J., & McCloy, R.A. (2004). Supervisor Ratings. In D. Knapp, R. McCloy, & T. Heffner (Eds.), *Validation of Measures Designed to Maximize 21st-Century Army NCO Performance* (Technical Report 1145). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Waugh, G.W. (2004). Situational judgment test. In D. Knapp, R. McCloy, & T. Heffner (Eds.), *Validation of Measures Designed to Maximize 21st-Century Army NCO Performance* (Technical Report 1145). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

White, L.A., (2002, October). *A Quasi-Ipsative Temperament Measure for Predicting Future NCO Leadership Performance.* Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

White, L.A., & Young, M.C. (1998, August). *Development and validation of the Assessment of Individual Motivation (AIM).* Paper presented at the annual meeting of the America Psychological Association, Washington, DC.

Young, M.C., Heggestad, E.D., Rumsey, M.G., & White, L.A. (2000, August). *Army Pre-implementation Research Findings on the Assessment of Individual Motivation (AIM).* Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

APPENDIX A

ASSESSING DIFFERENCES ACROSS ADMINISTRATION CONDITIONS

As described in the text of this report, a 2 x 2 design varied two factors: (a) instrument order for the LeadEx, SDI, and IQ-II, and (b) item order for these instruments. This design resulted in four instrument conditions. We examined two potential types of differences in instrument functioning across these conditions: (a) differences in the internal consistency reliability of instrument scales, and (b) differences in scale means.

Differences in Scale Reliabilities Across Conditions

To assess differences in scale reliabilities across administration conditions, we first computed a set of internal consistency reliability estimates for each instrument whose order of administration was varied during the data collection (i.e., the LeadEx, SDI, and IQ-II). A separate reliability coefficient was computed for each scale under each of four possible administration conditions (2 instrument orders x 2 item orders). Next, we conducted two sets of Feldt tests for the difference between two independent coefficient alphas (Feldt, 1969). These tests assessed whether reliability coefficients for each scale differed significantly across instrument orders or item orders, respectively. Specifically, to assess the significance of instrument-order effects, we conducted Feldt tests comparing (a) reliability coefficients for scales administered under different instrument orders *when one item ordering was used*, and (b) reliability coefficients for scales administered under different instrument orders *when the other item ordering was used*. To assess the significance of item order effects, we conducted Feldt tests comparing (a) reliability coefficients for scales with different item orders *when the first instrument order was used*, and (b) reliability coefficients for scales with different item orders *when the second instrument order was used*. Because we performed two tests of significance for assessing instrument order and item order effects for each scale, we adjusted the *p*-values associated with these tests using Bonferroni's correction procedure to maintain the experiment-wise Type I error rate at .05 (Hays, 1994). Table A.1 shows results of these analyses.

In general, minimal differences were found in the reliability of scales administered under different conditions. When significant differences were found, they tended not to be consistent across pay grade or across the administration factor controlled for in each test of significance. For example, although a significant difference was found in the reliability of the SDI Physical Conditioning scale between the first and second instrument orders, the difference was found only among E4 Soldiers when the item order A was used. Differences in this scale's reliability were not found for E5 Soldiers or when item order B was used. Given the prevalence of small effects and inconsistencies in the findings, it appears that the internal consistency reliability of the scales was not substantially affected by the ordering of instruments or items.

Table A.1. Internal Consistency Reliability Estimates by Administration Condition

	E4 Soldiers				E5 Soldiers			
	Instrument Order 1		Instrument Order 2		Instrument Order 1		Instrument Order 2	
Scale	Item Order A	Item Order B	Item Order A	Item Order B	Item Order A	Item Order B	Item Order A	Item Order B
LeadEx: 40-Item	.79[d]	.86[b,d]	.81	.80[b]	.78	.70[b]	.76	.82[b]
LeadEx: 24-Item	.69	.77	.68	.72	.72	.61	.70	.73
SDI : Dependability	.72	.64	.68	.57	.47	.59	.49	.57
SDI: Adjustment	.75	.71	.77	.74	.66[c]	.79[c]	.70	.75
SDI: Work Orientation	.76	.82[b]	.83[d]	.74[b,d]	.70	.80	.70	.73
SDI: Leadership	.80	.74	.81	.80	.76	.74	.78	.80
SDI: Agreeableness	.67	.70	.71	.65	.63	.67	.58	.66
SDI: Physical Conditioning	.59[a]	.71	.70[a]	.70	.55	.67	.60	.66
IQ-II: Tolerance for Ambiguity	.63[c]	.43[c]	.51	.49	.59	.48	.58	.49
IQ-II: Interpersonal Skills	.65[c]	.51[c]	.60	.47	.62	.53	.51	.47
IQ-II: Social Perceptiveness	.86	.87	.85	.84	.82	.82	.84	.83
IQ-II: Emergent Leadership	.86	.81	.83	.84	.79	.81	.81	.78
IQ-II: Manipulativeness	.71	.70	.74	.75	.69	.78	.73	.79
IQ-II: Hostility to Authority	.75	.75	.74	.71	.63	.70	.68	.62

Note. n_{E4} = 119-151 (per cell); n_{E5} = 73-92 (per cell). Instrument Order 1 = LeadEx, SDI, IQ-II; Instrument Order 2 = SDI, IQ-II, LeadEx. Item Order A = 1st half of items administered 1st, 2nd half of items administered 2nd; Item Order B = 2nd half of items administered 1st, 1st half of items administered 2nd. Statistically significant differences between alphas are noted as follows: [a]Significant difference between alphas for scales administered under different instrument orders within item order A; [b]Significant difference between alphas for scales administered under different instrument orders within item order B; [c]Significant difference between alphas for scales administered under different item orders within instrument order 1; [d]Significant difference between alphas for scales administered under different item orders within instrument order 2. For comparison of any two item or instrument orders on a single scale score, $p < .05$.

Differences in Scale Means Across Conditions

To assess the differences in scale means across administration conditions, we computed a two-way ANOVA (2 instrument orders x 2 item orders) for each scale. Table A.2 shows results of these analyses. Specifically, Table A.2 shows percentages of variance accounted for in the given scale by each factor in the administration design. Although some of the effects were statistically significant, they were small (i.e., only one of the significant effects accounted for more than 1.1% of the variance in scale scores).

Table A.2. Percentages of Variance Accounted for in Scale Scores by Administration Condition Factors

	E4 Soldiers			E5 Soldiers		
Scale/Factor	Test Order	Item Order	Test x Item	Test Order	Item Order	Test x Item
LeadEx: 40-Item	**0.9**	0.0	0.2	0.0	0.0	0.0
LeadEx: 24-Item	**1.0**	0.0	0.0	0.0	0.0	0.0
SDI : Dependability	0.0	0.2	**0.7**	0.0	0.0	0.2
SDI: Adjustment	0.3	0.3	0.0	**1.1**	0.4	0.0
SDI: Work Orientation	0.0	0.0	**0.6**	0.0	0.0	0.0
SDI: Leadership	**0.6**	0.0	0.3	0.0	**2.8**	0.3
SDI: Agreeableness	0.0	0.5	0.0	**1.0**	0.0	0.4
SDI: Physical Conditioning	0.0	0.0	0.0	0.0	0.0	0.0
IQ-II: Tolerance for Ambiguity	0.0	0.2	0.1	0.0	0.0	0.0
IQ-II: Interpersonal Skills	0.0	0.4	0.0	0.7	0.7	0.4
IQ-II: Social Perceptiveness	0.0	0.0	0.0	0.0	0.0	0.0
IQ-II: Emergent Leadership	0.0	0.4	0.0	0.0	0.0	0.0
IQ-II: Manipulativeness	0.3	**0.7**	0.0	0.3	0.0	0.0
IQ-II: Hostility to Authority	0.0	**1.0**	0.2	0.8	0.0	0.0

Note. $n_{E4} = 541$; $n_{E5} = 335$. Values in the table are ω^2 effect size statistics multiplied by 100. They are estimates of the percentage of variance accounted for in the given scale by the each of the factors (Hays, 1994). Bolded values indicate that the *F*-test for the factor was statistically significant ($p < .05$).

References

Feldt, L.S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika, 30*, 357-370.

Hays, W.L. (1994). *Statistics, Fifth Edition*. Fort Worth, TX: Harcourt Brace College Publishers.

APPENDIX B

NCO PROMOTION ANALYSIS SUPERVISOR WEBSITE INSTRUCTIONS FOR OBSERVED PERFORMANCE AND EXPECTED FUTURE PERFORMANCE RATINGS

Instructions for Performance Ratings

During this rating exercise you will rate your Soldier(s) on two types of scales:

1. Observed Performance - These rating scales were developed to assess current job performance. First, you will rate the Soldier's job performance in different areas. Then you will rate your Soldier's Overall Effectiveness and Senior NCO Potential.

2. Expected Future Performance - Each of these six rating scales take the form of a scenario that describes a major change predicted to occur in the future Army. After reading each scenario, you will rate how effectively you would expect the Soldier to meet these future NCO requirements.

Continue

Observed Performance Target Areas

You will be asked to rate your Soldiers on the following areas of NCO performance:

- MOS/Occupation-Specific Knowledge and Skill
- Common Task Knowledge and Skill
- Computer Skill
- Writing Skill
- Oral Communication Skill
- Level of Effort and Initiative on the Job
- Adaptability
- Self-Management and Self-Directed Learning Skill
- Demonstrating Integrity, Discipline, and Adherence to Army Procedures
- Acting as a Role Model
- Relating to and Supporting Peers
- Cultural Tolerance
- Selfless Service Orientation
- Leadership Skill
- Concern for Soldier Quality of Life
- Training Others
- Coordination of Multiple Units and Battlefield Functions
- Problem-Solving/Decision Making Skill
- Information Management

Continue

It is very important that you read and follow these directions carefully so that your ratings will be as accurate as possible.

For each performance area you will rate, the title is given in the gray box. A 7-point scale ranging from 1 (low) to 7 (high) appears under each rating area. Above the rating scale, statements are provided which describe different levels of performance effectiveness. For each Soldier you rate, you should first read these statements and decide which description most closely matches the Soldier's typical performance in that category. Try to think about how the Soldier usually performs. While everyone has "good days" and "bad days," base your ratings on how the Soldier performs most often.

In the example below, the rater is judging the performance area "Training Others." In this case, the rater gave the Soldier a rating of "5," indicating that the Soldier typically demonstrates behavior similar to the middle statement and occasionally shows some of the high-end behaviors in this area.

EXAMPLE

Target area

Training Others

How effectively does this Soldier provide relevant training experiences for subordinates?

Behavior descriptions

Is unaware of or ignores individual or unit training needs; fails to provide training experiences or gives subordinates inappropriate training; does not prepare well for formal training situations; fails to guide subordinates on technical training matters.	Usually ensures that important subordinate training needs are met when made aware of such needs; uses existing classroom or on-the-job training techniques; prepares as required for training sessions; sometimes guides and tutors subordinates on technical matters.	Actively seeks to be aware of individual or unit training needs; always makes time to provide relevant formal and informal training experiences for subordinates; prepares thoroughly for training sessions; effectively guides and tutors subordinates on technical matters.

Rating scale

LOW		MODERATE			HIGH	
1	2	3	4	5	6	7
○	○	○	○	◉	○	○

Cannot Rate ○

Before you make each rating, please read ALL the behavior description statements thoroughly so that you have a firm understanding of the kinds of behaviors that define different levels of effectiveness within each performance area.

Make your ratings by clicking on the button just below the appropriate number as shown above. Please do this for each of the scales. If you have not observed the Soldier's performance in this area and do not have a basis on which to judge the Soldier's performance, choose "Cannot Rate."

On every rating page, you can review the rating instructions by clicking on the "Review Instructions" link.

Things to Avoid

Before you begin your ratings we would like to alert you to some common mistakes raters make.

1. Everyone has strengths and weaknesses - your ratings should reflect this. Unconsciously, some raters let their general feelings (positive or negative) about a person influence their ratings. When this happens, they provide ratings that are higher or lower than deserved on all dimensions. Matching a Soldier's actual performance to the descriptions in the scales can help overcome this problem.

2. Some raters don't use the scales correctly. They may use mostly the high end or mostly the low end, regardless of who they are rating. Other raters just give ratings in the middle of the scale. Don't worry about trying to be nice or trying to show that you're tough to please. Match your Soldier's performance to the descriptions on the scale.

3. Some raters are overly influenced by a recent event and base their ratings too heavily on the last thing they saw the Soldier do. As you make your ratings, think about the Soldier's performance during the whole time you have supervised or worked with the Soldier.

Things to Remember

- Most Soldiers do not perform at the same level in all areas. Most often, a Soldier has some strong areas and some areas where he/she needs improvement. Your ratings should accurately reflect your Soldier's strengths and weaknesses.

- Making accurate ratings is the key to success. While you should keep the common mistakes in mind, if your Soldier always performs at the highest level (or the lowest) your ratings should reflect that.

- Most importantly, using the scales keeps all raters "on the same page" and ensures that all Soldiers are measured objectively.

- **Remember, these ratings are for research purposes only and cannot help or hurt your Soldier.**

Continue

Instructions for Expected Future Performance Ratings

You will be given descriptions of the major conditions predicted to be characteristic of the future Army. After you read each description, please rate how effectively you would expect the Soldier you are rating to meet those future requirements.

Continue

Expected Performance Under Future Army Conditions
Scenario #1: Increased Requirements for Self-Direction and Self-Management

The predicted changes in missions, technology, structure, and tactics will require that NCOs have a greater ability to guide their own professional development and manage their personal affairs (e.g., Family concerns and financial matters). Increasing mission diversity and frequency will be disruptive. For example, frequent deployments away from U.S. home bases will require a strong ability to manage personal matters effectively. In addition, the restructuring of the Army into smaller, more independent units will require that NCOs have a greater ability to take initiative in their actions and make their own decisions without direct supervision. Finally, due to greater technological change and more frequent changes in missions, there is an expectation that individual NCOs will need to assume more and more responsibility for their own training. That is, they will be required to identify their own training needs and to seek out training experiences that meet these needs. They will need to evaluate their own training accomplishments and take corrective steps if necessary.

1. How effectively would you expect the Soldier to meet these future NCO requirements?

Not likely to meet the NCO demands described under these conditions.		Likely to be generally successful, but will struggle to meet the NCO demands described under these conditions.			Likely to successfully meet or exceed NCO demands described under these conditions.	
LOW		**MODERATE**			**HIGH**	
1	2	3	4	5	6	7
○	○	○	○	○	○	○

Next Rating Item